## DOCUMENT RESUME

ED 147 722 CG 012 058

| | |
|---|---|
| TITLE | Evaluation and the School Counselor. |
| INSTITUTION | National Education Association, Washington, D.C.<br>National Association of School Counselors. |
| PUB DATE | 77 |
| NOTE | 115p. |
| AVAILABLE FROM | National Association of School Counselors, 1201 16th St. N.W., Washington, D.C., 20036 ($5.00 each) |
| EDRS PRICE | MF-$0.83 Plus Postage. HC Not Available from EDRS. |
| DESCRIPTORS | *Accountability; *Counselor Evaluation; *Counselors; *Evaluation; Guides; *Management by Objectives; *Professional Associations; Program Development |

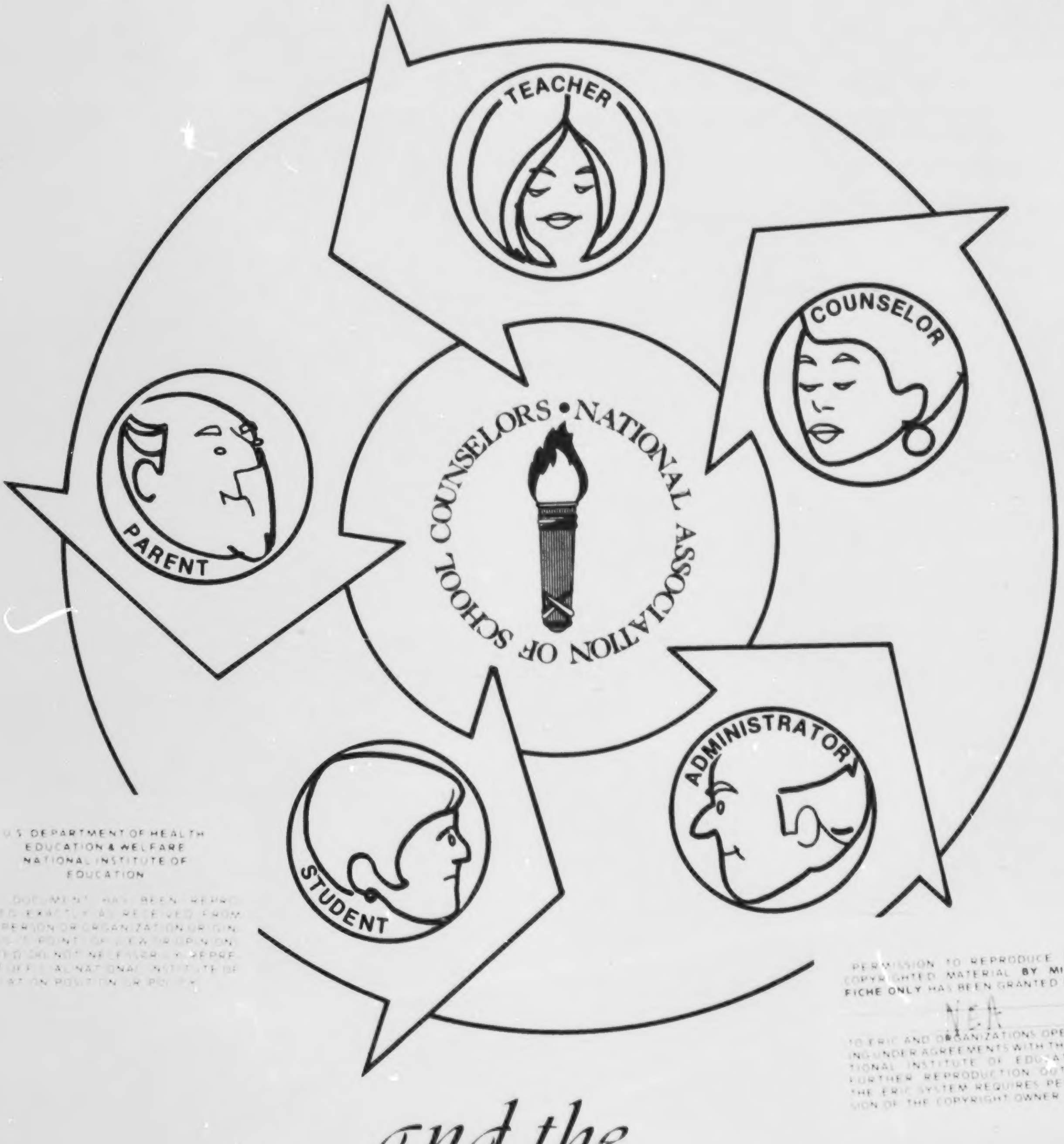
ABSTRACT

This guide, prepared for use by education associations involved in the process of developing positions, training programs, and agreements on evaluation, discusses many aspects of the evaluation procedure. Included are sections on the counselor's role, the requisites for evaluation, the procedures, criteria, evaluating instruments, etc. There is also a statement of the National Association of School Counselors' on evaluation. In addition there is a description and discussion of management-by-objectives, plus a lengthy section on accountability and its various aspects. (PFS)

***********************************************************************
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
***********************************************************************

ED147722

# Evaluation


TEACHER
COUNSELOR
PARENT
NATIONAL ASSOCIATION OF SCHOOL COUNSELORS
STUDENT
ADMINISTRATOR

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL **BY MICROFICHE ONLY** HAS BEEN GRANTED BY

N.E.A.

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

CG012058

# and the School Counselor

95-

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

# TABLE OF CONTENTS

PAGE

## THE COUNSELOR AND EVALUATION

Introduction 1
Instructions 3
The Counselor's Role in Evaluation Process 4
The Requisites for Evaluation 5
The Procedures for Evaluation 7
Due Process in Evaluation Procedures 10
Criteria for Evaluation 12
Evaluators and Their Roles 14
The Evaluating Instrument 15
Evaluation for Professional Development 16
Additional Evaluation Processes 20
Instrument for Evaluating Administrators 22
Administrative Planning 23
Relations With Subordinates 24

## NASC POSITION PAPER

Evaluation and Subjective Ratings 25
Position Recapitulation 25
The Issues 26

## IS MBO THE WAY TO GO?

An Educator's Guide to Management By Objectives 27
What Is MBO? 28
The Cash Register and the Curriculum 31
Managers and Management 34
The Technology of Management 37
Some Objectives to Objectives 39
The Teacher and MBO 42
The Local Association and MBO 47
References 52

## ACCOUNTABILITY

Some Manifestations of Accountability 53
Problems in Using Pupil Outcomes for Teacher Evaluation 54
Roles and Responsibilities of Groups Concerned with Student Evaluation Systems 73
Why Should All Those Students Take All Those Tests? 85
Stages in Developing An Evaluation Program 94
Criteria for Evaluating State Education Accountability Systems 104

# THE COUNSELOR AND EVALUATION

## INTRODUCTION

Most evaluations now conducted are a negative, threatening process with the primary purpose of making employment decisions -- whether counselors whould be hired, fired, placed on tenure, or promoted. Research has shown that when evaluation is conducted for such purposes there is little likelihood that it will result in imporved counseling.

With the clamor for increased school accountability there has been a concomitant interest in evaluation. The negative connotations of evaluation have been intensified by attempts to measure performance on the basis of student performance, frequently through the use of student scores on standardized tests.

Despite these difficulties, evaluation should not be eliminated; rather, it should be reformulated so that its primary purpose becomes that of helping counselors do a better job. It should take place in a nonthreatening atmosphere with an understanding that the improvement of a counselor's performance is a cooperative effort between the counselor and the evaluators. The instructional improvement purpose can be facilitated if counselors are involved in all phases of evaluation -- in establishing the criteria, in developing the procedures, and in conducting the evaluation.

The purposes of evaluation will also be more readily attained if provisions are made for special preparation and training for those people who are involved in evaluating counselors. Rather than assuming that someone with administrative experience is qualified to evaluate counselors, evaluation must be recognized as a complex task requiring special abilities and skills.

In addition, evaluation should encourage creativity and divergent thinking rather than conformity to a set pattern or model of counseling behavior. This can best be achieved when more than one person is involved in the evaluation process. A team evaluation system based on feedback and consultation with supervisors, fellow counselors, teachers, students, and self-evaluation should be used to provide data for making decisions about improvement. Finally, the appraisal of counselor performance should be a continuing process which results in individualized inservice programs designed to fit the unique needs identified in the evaluation.

The association can accomplish several significant purposes by advocating and attaining good evaluation procedures and processes for its members.

First, it can provide protection of counselors against restrictive, arbitrary, or punitive evaluation. It can also provide a positive image to the public -- the consumers of education -- by demonstrating that counselors do wish to be held accountable for those decisions and performances over which they have control. Finally, and most important, association attainment of sound constructive evaluation procedures can help each member experience the satisfaction which accompanies constructive improvement of their professional performance.

This guide was prepared for use by associations involved in the process of developing positions, training programs, and agreements on evaluation. Its objectives are four-fold:

1. To create an awareness on the part of the association of potential means whereby counselors can participate in the process of their evaluation;
2. To suggest a systematic procedure for achieving meaningful input by the association in the process of negotiating evaluation procedures;
3. To provide samples of what might be achieved through illustrated contract language and evaluation devices;
4. To present a substantive base and means for training programs whereby we can more effectively participate in the evaluation programs.

The guide is intended as a working document to be used by associations selecting from, adding to, and adapting the materials to fit their own circumstances and needs.

# INSTRUCTIONS

The guide identifies eight levels of potential involvement by the association in the process of developing and implementing effective evaluation systems in school districts.

These levels are listed at the top of the page under the heading "Does your association want to negotiate?" followed by instructions to indicate "yes" or "no". It should be noted that the level of involvement is the prerogative of the association, thus if "no" is checked for any level, turn the page and proceed to the next level. However, if "yes" is checked, move downward on the page to the section titled "Considerations: Have you Weighed the importance of Including in the Contract?" This section is designed to be used as a checklist for priority determination. Blank spaces have been provided in each level for the addition of considerations considered important by the user.

Contract language is provided for implementing each level of involvement. This language should be considered illustrative of the considerations presented. Negotiators should revise as appropriate.

## LEVEL I: THE COUNSELOR'S ROLE IN THE EVALUATION PROCESS

Does your association want to negotiate a meaningful role in the evaluation process?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level II.

---

Considerations:

Have You Weighed the Importance of Including in the Contract?

1. Evaluation Committee?
   - a. that counselors and teachers constitute a majority of the committee membership?
   - b. that all be members of the Association?
   - c. that the members on the committee be appointed by the Association?

2. The functions of the Evaluation Committee?
   - a. to design the instruments/procedures utilized?
   - b. to select the instruments utilized?
   - c. to evaluate potential instruments/procedures?
   - d. to approve any instruments/procedures utilized in the evaluation process?

3. The meeting and reporting dates of the Evaluation Committee?

4. ____________________________________________________________

____________________________________________________________

Sample Contract Language:

1. Prior to September 15, the parties to the Agreement shall establish an Evaluation Committee which shall consist of three (3) representatives of the Board and four (4) representatives appointed from the Association by the President, who shall preside as chairperson of the committee.

2. The Evaluation Committee shall develop specific criteria and procedures for the evaluation of certificated staff incorporated by the Agreement.

3. The recommended criteria and procedures shall be submitted for adoption by the parties not more than 45 days from the establishment of the Committee.

4. No evaluations of certificated staff shall be conducted prior to adoption and implementation of agreed upon criteria and procedures.

5. All agreed upon criteria and procedures shall become an addendum to this Agreement.

## LEVEL II: THE REQUISITES FOR EVALUATION

Does your association want to negotiate the requisites for evaluation?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level III)

---

Considerations:

Have you Weighed the Importance of Incuding in the Contract:

1. That the primary objective of evaluation is to improve the counseling services?

2. That evaluation observations should be conducted openly and with the full knowledge of the counselor?

3. That mechanical observation devices should be used only with the consent of the counselor observed?

4. That all personnel involved must be well informed about the evaluation system and that the purposes, the criteria, and the respective roles of those involved must be communicated to all concerned?

5. That evaluation procedures are most effective when the personnel being evaluated are actively involved in the process? And that a pre-evaluation conference is necessary in order that the observer can be made aware of the lesson's context, objectives and methods?

6. That since effective evaluation is a continuous process, observations must be made on numerous occasions over an extended period of time?

7. That each observation should be for an appropriate amount of time, typically the length of the specific instructional activity?

8. That judgments, ratings, and recommendations must be made in a manner to minimize feelings of insecurity by both the evaluators and those being evaluated?

9. That each must be afforded the opportunity to communicate their rationale for utilizing various instructional methodologies?

10. That if we are to improve our professional compentencies, we must be assured the opportunity to receive and react to observations, judgments, and recommendations; therefore a

conference to communicate the evaluator's findings and recommendations must be an integral part of the evaluation process?

11. That evaluations must be made in good faith as a means to bring about improved insturction and not to collect information to support preconceived judgments?

12. That the criteria for evaluation must be identified in the evaluation instrument?

13. That we should be encouraged to use evaluations of various types to supplement those of supervisors/administrators?

14. ______________________________________________

______________________________________________

Sample Contract Language:

1. The primary objective of evaluation shall be the improvement of counseling services.

2. All monitoring or observation of the work performance of a staff member will be conducted openly and with the full knowledge of the staff member. The use of eavesdropping, public address or audio systems, and similar surveillance devices will be strictly prohibited.

3. All methods of observation shall be used only with prior knowledge by and consent of said staff member being evaluated.

4. The criteria and procedures shall be applied uniformly throughout the District and any deviations from them shall constitute a null and void evaluation and a direct violation of the Agreement.

5. Prior to implementation of the criteria and procedures for evaluation a workshop shall be conducted by the administration and the Association to fully inform its members of said final evaluation criteria and procedures.

6. All evaluations shall be conducted in good faith for the purpose of improving services and in accordance with the provisions of this Agreement.

## LEVEL III: THE PROCEDURES FOR EVALUATION

Does your association want to negotiate the procedures (steps) to be implemented in the teacher evaluation process?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level IV)

---

Considerations:

Have You Weighed the Importance of Including in the Contract:

1. The category (probationary or tenure) of the employees evaluated?
2. The frequency of evaluations?
3. The length of the evaluation observations?
4. The dates of the evaluations?
5. The necessity of pre-evaluation conferences?
6. The necessity of post-evaluation conferences?
7. The necessity of formally notifying employees of the evaluation findings?
8. The necessity of assuring employees an opportunity to react to the evaluations?
9. The necessity of acknowledging the members strengths?
10. The necessity of apprising faculty of the evaluation process?
11. The necessity of specifying suggestions for improvement into the evaluation instrument?
12. ____________________________________________

    ____________________________________________

Sample Contract Language:

1. A probationary employee shall be evaluated not less than twice per semester. The initial evaluation shall be preceded by a period of not less than thirty (30) working days prior to the evaluation. There shall be a fifteen (15) work-day period between each evaluation unless otherwise requested by the employee.

2. Tenured employees shall be evaluated not less than once each semester.

3. Each evaluation shall be conducted with not less than two (2) periods of sessions by the evaluator. Said periods shall be attended within three (3) class days of the initial observation unless otherwise agreed to by the employee involved.

4. Each evaluation shall be initiated by a pre-evaluation conference, between the evaluator and the employee to be evaluated, that shall include a review of the evaluation criteria and procedures, discussion of the expectations of the parties, and an agreement on the dates and specific times for the observations.

5. A copy of each written evaluation shall be given the employee and a conference shall be scheduled between the parties not more than five (5) days following the final classroom observation. Should the employee feel that the evaluation is incomplete, inaccurate, or unjust, the employee may submit the objections in writing and they shall become an official addendum to the written evaluation. A copy signed by the parties shall be given to said employee. A signature by the employee shall not attest to the accuracy and/or validity of the written evaluation.

6. The written evaluation shall contain acknowledgement of the employee's strengths. Should deficiencies be recorded, the evaluator shall provide the employee with specific written recommendations for improving the quality of insturction and for eliminating any deficiencies noted in the written evaluation.

7. The building principal or immediate superviosr shall provide the employee with definite, positive assistance to improve the quality of work. If the evaluator finds an employee lacking, the reasons therefore shall be set forth in specific terms, as shall an identification of the specific ways in which the employee is to improve and of assistance to be given by the principal or supervisor. If the principal or supervisor cannot provide the necessary assistance, appropriate assistance shall be provided by other supervisory, consultant, or administrative personnel.

   In subsequent evaluation reports, failure to again note a specific deficiency shall be interpreted to mean that adequate improvement has taken place.

8. No material originating after original employment shall be placed in an employee's personnel file unless the employee has had an opportunity to review the material. The employee may submit a written notation regarding any material and the same shall be attached to the file copy of the material in question. If the employee believes that material to be placed in his file is inappropriate or in error, they may receive adjustment through the grievance procedure, whereupon the material shall be corrected or expunged from the file. If the employee is asked to sign material placed in their file, such signature shall be understood to indicate their awareness of the material but in no instance shall said signature be interpreted to mean agreement with the content of the material.

9. Any complaint made against an employee or person for whom the employee is administratively responsible, by any parent, student, or other person, shall promptly be called to the attention of the employee if said complaint is to be placed in the employee's personnel file. The procedures of Section 6 of this clause apply to these records as well as all others that are placed in the file.

## LEVEL IV: DUE PROCESS IN EVALUATION PROCEDURES

Does your association want to negotiate due process for the evaluation procedures?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level V)

---

**Considerations:**

Have you Weighed the Importance of Including in the Contract:

1. Assurance of prior consent as to when evaluations are to occur?
2. The requirement that all observations will be succeeded by a follow-up conference?
3. The prerequisite that a written copy of the evaluation must be given to the employee.
4. That the employee will be assured the cooportunity to respond to the evaluation in writing within a prescribed time? And that this response will be placed in the employee's personnel file?
5. That the employee is assured the right and process of appeal should he/she be dissatisfied with the evaluation?
6. That the employee be guaranteed free and open access to his/her personnel file to all items excepting those rendered in confidence such as letters of recommendation and placement papers?
7. That persons having access to the evaluation information and the conditions under which they are entitled to this information should be prescribed by the contract?
8. ________________________________________

   ________________________________________

**Sample Contract Language:**

1. No evaluation shall occur without the prior consent of the employee.
2. Each employee, upon his employment, or at the beginning of the school year, whichever is later, shall be apprised of the specific criteria upon which he will be evaluated. No criterion shall be used which has not previously been approved by the Association as applicable for the position held by the employee being evaluated.

3. Each evaluation shall be succeeded by a post-evaluation conference.

4. Each employee shall receive a written copy of his/her evaluation prior to the prescribed post-evaluation conference.

5. Should the employee raise objections to said evaluation, such employee shall have recourse via all applicable provisions of this Agreement without limitations.

6. Each employee shall have his/her evaluation placed within his/hre active personnel file and shall, upon verbal request, have the right to review the contents within. Upon request of the employee, a respresentative of the Association may accompany said employee in this review.

7. No evaluation materials relevant to discipline or reemployment may be placed in the employee's personnel file without first giving the employee a copy and an opportunity for placing a written rebuttal to such material in the file.

8. All communications obtained by an employee in the course of his professional duties and deemed by said employee to be of a confidential nature shall not, except with his consent, be disclosed to anyone unless said disclosure is required by law.

9. No individual shall have access to any employee's personnel file, regardless of position or office, unless such access is for reason which is demonstrably relevant and significant.

# LEVEL V: THE CRITERIA FOR EVALUATION

Does your Association want to negotiate the nature of the criteria to be utilized in the evaluation?

_____ Yes (If yes, proceed to Considerations listed below)

_____ No (If no, proceed to Level VI)

---

## Considerations:

Have You Weighed the Importance of Including in the Contract:

1. The evaluations must be based on identified and pre-determined criteria?

2. The evaluations must be based upon observed employee behavior relevant to the education process?

3. That only those professional criteria which are demonstrably relevant to the assignment as opposed to personal criteria and life style, should be included in the evaluation instrument?

4. That criteria should focus on the relevance of educating style, diagnosis of student needs, prescription of appropriate programs, student interaction and the degree to which student behavior is changed?

5. That both the criteria and the suggestions for improvement must be rendered in a form that communicates effectively to the employee.

6. ____________________________________________

____________________________________________

## Sample Contract Language:

1. Evaluation shall be on identified and pre-determined criteria which are based on:

   a. Observed employee behavior relevant to the educational process;

   b. Professional characteristics which are demonstrably relevant to productivity;

   c. Relevance of educational style, diagnosis of student needs, prescription of appropriate programs, student interaction and the degree to which student behavior is changed.

2. Evaluation responses not relevant to the above shall be expunged from said evaluation prior to placement in the employee's personnel file.

3. Any abridgement of professional and civil rights protected by federal and state constitutions, applied law, case law, and the Agreement shall not be considered acceptable entries in the written evaluation and shall be expunged from the evaluation placed in the personnel file.

## LEVEL VI: THE EVALUATORS AND THEIR ROLES

Does your association want to negotiate the evaluators and their respective roles?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level VII)

---

Considerations:

Have You Weighed the Importance of Including in the Contract:

1. That the purpose(s) of the evaluation must be identified?

2. That the evaluators must be identified by title or position?

3. That those with evaluation responsibilities must have had formal training in the evaluation process?

4. ____________________________________________________________

____________________________________________________________

Sample Contract Language:

1. Personnel conducting evaluations shall submit in writing to the committee a list of credentials and qualifications which acknowledge the skills and proficiencies of said individual's ability to evaluate within the spirit of this Article. Should the committee denote a deficiency within an evaluator, recommendations shall be made to qualify said person. Should an evaluator be judged deficient by the committee, provisions shall be made by such to provide proper evaluation for the employees involved.

2. The evaluation of each employee shall be performed by___________ (principal, assistant principal, department head, etc.)

# LEVEL VII: THE EVALUATION INSTRUMENT

Does your association want to negotiate the form of the instrument(s) to be utilized?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No (If no, proceed to Level VIII)

---

## Considerations:

Have You Weighed the Importance of Including the Contract:

1. The criteria for selection of the evaluation forms used?
2. A copy of an approved evaluation form that has been designed, revised, and/or adapted for local use?

## Sample Contract Language:

Following are two alternative types of evaluation forms to be used. The first is, in essence, a blank sheet of paper with the entries to be made on it specified by contract. The second is a definite rating scale instrument with criteria and sections to be checked and entered.

### SAMPLE I.

1. The evaluation form shall be applied equally to all affected personnel within the district.
2. The evaluation form shall contain the following:
   a. Name of evaluator
   b. Title or position of evaluator
   c. Employee being evaluated
   d. Date(s)
   e. School
   f. Site
   g. Work assignment in class/subject/counseling
   h. Written observations of the evaluator which include first-hand knowledge of strengths, weaknesses, and remediation as applicable. All recommendations shall be supported by observed first-hand knowledge.

SAMPLE II.

## EVALUATION FOR PROFESSIONAL DEVELOPMENT

I. Name____________________ School__________________ Date__________

Tenured____________ Probationary_______________

II. Professional Activities Observed:

| Professional Assignment | Date-Time-Place |
|---|---|
| __________________________ | ______________________________ |

III. Evaluator:_________________________ Position________________

IV. Performance Areas and Ratings:

Explanation: Place an "X" in the space deemed most appropriate. A four point scale is used in which 4 is outstanding; 3 is satisfactory; 2 needs improving; and 1 is unsatisfactory.

| A. Preparational Competencies: | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| The performance of the employee observed indicates: | | | | |
| 1. knowledge of job assignment | ____ | ____ | ____ | ____ |
| 2. an interest and enthusiasm for the job assignment | ____ | ____ | ____ | ____ |
| 3. an understanding of the students | ____ | ____ | ____ | ____ |
| 4. a genuine concern for the welfare of students | ____ | ____ | ____ | ____ |
| 5. a knowledge of current developments in the assigned area | ____ | ____ | ____ | ____ |
| 6. other (list)____________<br>____________________<br>____________________ | ____ | ____ | ____ | ____ |

Observer remarks and/or suggestions for improvement:__________________

________________________________________________________

| B. Instructional Skills and Techniques | | | | |
|---|---|---|---|---|
| The performance of the counselor and/or the behavior of students indicates effective: | | | | |
| 1. planning for counseling program | ____ | ____ | ____ | ____ |
| 2. organization of counseling activity | ____ | ____ | ____ | ____ |
| 3. ability to motivate students | ____ | ____ | ____ | ____ |
| 4. ability to maintain student interest | ____ | ____ | ____ | ____ |

| Instructional Skills & Techniques (continued) | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| 5. use of varied methodologies | | | | |
| 6. student participation | | | | |
| 7. assessment techniques | | | | |
| 8. assignment making skills | | | | |
| 9. selection and utilization of materials | | | | |
| 10. student achievement | | | | |
| 11. acceptance of student points of view | | | | |
| 12. poise and confidence by the counselor | | | | |
| 13. other (list)________________ | | | | |
| ______________________________ | | | | |

Observer remarks and/or suggestions for improvement:_______________

______________________________________________________________

______________________________________________________________

C. Management Skills

The performance of the counselor, the behavior of students observed, and the counseling setting reflect effective:

| | | | | |
|---|---|---|---|---|
| 1. pupil control | | | | |
| 2. enforcement of school policies | | | | |
| 3. maintenance of school reports and records | | | | |
| 4. maintenance of a counseling setting conducive to effective counseling | | | | |
| 5. other (list) ________________ | | | | |
| ______________________________ | | | | |

Observer remarks and/or suggestions for improvement:_______________

______________________________________________________________

______________________________________________________________

D. Professional Characteristics

The behaviors of the counselor observed typically reflect:

| | | | | |
|---|---|---|---|---|
| 1. effective ability to relate to students | | | | |
| 2. maturity of professional judgment | | | | |
| 3. an ability to relate to parents, fellow teachers & administrators | | | | |
| 4. active support for school policy | | | | |

| Professional Characteristics (continued) | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| 5. a willingness to accept criticism | ____ | ____ | ____ | ____ |
| 6. receptiveness to new ideas | ____ | ____ | ____ | ____ |
| 7. willingness to work cooperatively with others | ____ | ____ | ____ | ____ |
| 8. effective communication skills | ____ | ____ | ____ | ____ |
| 9. punctuality and dependability | ____ | ____ | ____ | ____ |
| 10. good mental and physical health | ____ | ____ | ____ | ____ |
| 11. appropriate leadership capability | ____ | ____ | ____ | ____ |
| 12. effective participation in curricular planning activities | ____ | ____ | ____ | ____ |
| 13. effective participation in the process of resolving school problems | ____ | ____ | ____ | ____ |
| 14. other (list)________________<br>________________________<br>________________________ | ____ | ____ | ____ | ____ |

Observer remarks and/or suggestions for improvement:____________

________________________________________________

________________________________________________

________________________________________________

V. Summary of Major Strengths:

## VI. Summary of Instructional Concerns with Specific Suggestions for Improvement:

## LEVEL VIII: ADDITIONAL EVALUATION PROCESSES

Does your association want to negotiate additional items?

_____ Yes (If yes, proceed to the Considerations listed below)

_____ No

---

**Considerations:**

Have You Weighed the Importance of Including in the Contract:

1. That all professional employees shall be encouraged to utilize evaluation procedures, other than the traditional supervisor ratings, in the form of colleague (peer) student, parental and self-evaluations for the purpose of assuring comprehensive evaluations.
2. That all administrators shall be evaluated by their staff at least twice per year.
3. ______________________________________________

**Sample Contract Language:**

1. Evaluation of Administrators -
    - a. All administrators shall be evaluated by the professional staff they supervise not less than once per year.
    - b. The evaluation of an administrator shall be coordinated by a committee of the professional staff he/she supervises, selected by all the professional staff so supervised. This committee shall oversee the process and reduce the evaluation results to writing.
    - c. Within two weeks of having analyzed the evaluations, the Committee shall meet with the administrator being evaluated and advise him/her as to improvements which are expected in their professional services.
    - d. If the administrator fails to make improvements suggested by the Committee, the professional staff may so advise the Superintendent. The Superintendent shall then meet jointly with representatives of the professional staff and the administrator.
    - e. The Board, upon request of the Association following the action outlined in Section 'd' above, shall meet with the Association to discuss the situation and seek resolution of differences. Said meeting shall be open or closed at the request of the Association.

2. Employee Self-evaluation -

The Board and the Association agree that providing a variety of means for employees to appraise their own professional behavior in systematic ways is important in improving the quality of instruction. The Board and the Association also agree that the formal system of evaluation now in use serves the needs of the Board to discharge its responsibility for assurance of high-quality instructional services.

The Board and the Association recognize, however, that there is significant value in providing employees with a variety of means to pursue to the fullest their personal desire to engage in self-stimulated, private, secure, and systematic examination of their own educational practices in an effort to continually improve the quality of their services.

The Board agrees, therefore, to provide employees with a variety of means to examine their own teaching behavior in privacy, with full and complete assurance that the results of such appraisal shall be controlled by the individual employee. It agrees further that there shall be no carry-over into the system of evaluation used by the ____________School District for purposes of making decisions about retention, promotion, and dismissal.

For the purpose of distinguishing between the two processes, the effort to assist employees with examination of their own practices shall be designated the Self-Appraisal Program. The District-operated evaluation process shall continue to be known as the Employee Evaluation Program.

To carry out the purposes of this understanding, the Board agrees to:

a. Recognize a Committee appointed by the Association as the agency to manage the Self-Appraisal Program.

b. Insist that all employees honor the security requirements of the Self-Appraisal Program. Such security shall be maintained at the discretion of the individual employee whose practices are under appraisal.

c. Provide the above-mentioned Committee an amount up to __________ dollars ($ ) for actual and necessary expenses and assigned time to examine the extant self-appraisal systems and identify an appropriate number of such systems for use in the District.

d. Provide the financial and physical resources necessary to train a cadre of volunteer employees in the use of the self-appraisal systems selected by the Committee to the extent that said cadre can in turn train other employees who request such training.

e. Provide assigned time for volunteers to become trained in the use of their choice among the selected systems, such training to be given by a member of the training cadre mentioned above.

# INSTRUMENT FOR EVALUATING ADMINISTRATORS

## THE PURDUE RATING SCALE FOR ADMINISTRATORS AND EXECUTIVES

H. H. Remmers and R. L. Hobson

Instructions:

You are asked to rate your administrator on the following scale. Your ratings will be anonymous -- the administrator will never know how you personally appraised him/her; they will receive only the average responses of all those who rate him/her. Read the items carefully. Decide which of the five possibilities best describes your administrator. Place the number corresponding to your choice in the blank at the right.

Name of person being rated ______________________________

### I. INTELLECTUAL BALANCE

1. Possesses general knowledge:
(5) Very broad (4) Fairly broad (3) Limited (2) Very limited (1) Lacking ____ 1

2. Possesses specific knowledge in his own field:
(5) Up-to-date (4) Good (3) Fair (2) Poor (1) Lacking ____ 2

### II. EMOTIONAL BALANCE

3. Is emotionally poised and calm:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never ____ 3

4. Has adequate self-confidence:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never ____ 4

5. Is concerned with his own personal problems:
(5) Never (4) Seldom (3) Sometimes (2) Usually (1) Always ____ 5

6. Welcomes differences inviewpoint:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never ____ 6

### III. ADMINISTRATIVE LEADERSHIP

7. Welds staff into a unit with clearly recognized goals:
(5) Exceptionally well (4) Very well (3) Quite well (2) Poorly (1) Very poorly ____ 7

8. Uses democratic procedures wherever possible:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never ____ 8

9. Inspires subordinates to independent creative work:
(5) Always (4) Sometimes (3) Seldom (2) Never (1) Makes creative work repulsive ____ 9

## IV. ADMINISTRATIVE PLANNING

10. Makes plans carefully and adequately:
(5) Always (4) Usually (3) Sometimes (2) Seldom
(1) Never _____ 10

11. Is alert to recognize or devise useful innovations:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 11

12. Understands the objectives and interrelationships of his entire work:
(5) Exceptionally well (4) Very Well (3) Quite well
(2) Poorly (1) Very poorly _____ 12

13. Does a good job of systematizing and coordinating units of work:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 13

14. Has knowledge of pertinent details of his subordinates' work:
(5) Very good (4) Good (3) Fair (2) Poor (1) Not at all _____ 14

## V. USE OF FUNDS

15. Employs as capable personnel as possible:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 15

16. Selects equipment wisely:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 16

17. Makes effective effort to obtain funds for self-improvement of subordinates:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 17

## VI. CAPACITY FOR WORK

18. Works hard:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 18

19. Welcomes additional responsibilities:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 19

20. Meets emergencies in his work competently:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 20

## VII. ACCOMPLISHMENT

21. Conducts his work as expeditiously as possible:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 21

22. The essential work of his organization get done on time:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 22

23. The important work of his organization is completed:
(5) All of it (4) Most (3) Some (2) Little (1) None _____ 23

## VIII. RELATIONS WITH SUBORDINATES

24. Compliments and thanks his subordinates appropriately
and sincerely:
(5) Very frequently (4) Quite frequently (3) Sometimes
(2) Seldom (1) Often criticizes negatively _____ 24

25. Is available to counsel and assist subordinates:
(5) Sufficiently (4) Almost sufficiently (3) Sometimes
(2) Seldom (1) Never _____ 25

26. Recognizes and rewards meritorious achievement of his
subordinates:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 26

27. Possesses insight into the problems encountered by his
subordinates:
(5) Complete (4) Much (3) Some (2) Little (1) None _____ 27

28. Is honest and dependable in dealings with subordinates:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 28

29. Displays unwarranted favoritism to some subordinates:
(5) Never (4) Seldom (3) Sometimes (2) Often (1) Continuously _____ 29

30. Appropriates ideas and work of subordinates to improve own
understanding:
(5) Never (4) Seldom (3) Sometimes (2) Often (1) Continuously _____ 30

31. Does everything possible, consistent with a subordinate's ability
and achievement to advance him/her:
(5) Always (4) Usually (3) Seldom (2) Never
(1) Curbs advancement _____ 31

32. Is just and considerate in discharging subordinates:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 32

33. The general morale of his staff:
(5) Exceptionally high (4) Good (3) Fair (2) Poor (1) Very low _____ 33

## IX. PUBLIC RELATIONS

34. Promotes public relations:
(5) Actively good (4) Fair (3) Poor (2) Not at all
(1) Actively bad _____ 34

35. Attempts to orient his work to the welfare of society at large:
(5) Exceptionally well (4) Well (3) Fairly well (2) Indifferently
(1) Poorly _____ 35

36. Team work; conforms to the purposes and plans of the organization
which he serves; does not seek unfair advantage for his unit:
(5) Always (4) Usually (3) Sometimes (2) Seldom (1) Never _____ 36

Copyright: THE PURDUE RESEARCH FOUNDATION, Purdue University, Lafayette, Indiana

# NASC POSITION PAPER

## EVALUATION/RATING OF PROFESSIONALS

### EVALUATION AND SUBJECTIVE RATINGS

The Association believes that it is a major responsibility of educators to participate in the evaluation of the quality of their services.

Evaluation should be positive and the concept of evaluation should be directed toward the improvement of instruction and professional growth.

The Association insists that each educator must have immediate access to and copies of all items in his personnel file (except privileged communications related to his initial employment) and shall have the right to attach a written response to any item. A procedure shall be established to remove inappropriate or unfounded material from personnel files. A copy of any evaluation report placed in the educator's file must be given to the educator.

The Association asserts that a professional employee may at any time review the contents of his personnel file and receive copies of any documents contained therein at his/her request. In examining his/her file, an employee shall be entitled to be accompanied by an association representative to examine his/her file in his/her behalf. An employee shall be notified in writing of any document placed in his/her personnel file on the date of its placement with the right to annotate such documents. All personnel files shall be kept in one location.

The Association opposes the use of subjective methods of evaluating professional performance for the purpose of setting salaries.

The Association believes that a single state mandated rating form has advantages over a proliferation of local forms that may endanger the rights of professional employees.

Local forms for rating or evaluation must be developed with active participation of the local association, have the approval of the local association, and must have the safeguards of the state forms.

### POSITION RECAPITULATION

This position speaks to:

A. The responsibility of educators to participate in the evaluation of the quality of their services.

B. Making evaluation positive and directed toward improvement of instruction and professional growth.

C. The right of educators to have access to personnel file items, attach written responses, removal of inappropriate materials.

D. Opposition to subjective ratings.

E. The advantages of a single mandated rating form over a proliferation of local forms.

F. The local forms being developed with the association having active participation and approval by that association.

## THE ISSUES

The increasing number of suspensions has sensitized educators and association advocates to the issues of evaluation/rating. Our resolutions call for a positive process to improve services. In practice we see the process being manipulated to eliminate counselors from their positions. This is happening in a majority of areas of the country. Counselors none the less are becoming more defensive and hostile to the process.

Two basic issues confront us:

1. Protection of the due process rights of all educators (our members).
2. Reducing the anxiety level of educators so that the evaluation process does have the positive impact of improving instruction called for in our resolutions.

# IS MBO THE WAY TO GO?

## *AN EDUCATOR'S GUIDE TO MANAGEMENT BY OBJECTIVES**

Management by objectives (MBO) is one way to run a business. For the past twenty years it has been tried, refined, and sometimes found useful in factories, in processing plants, and in sales organizations.

More recently, and much more quietly, the U. S. Office of Education (USOE) has supported such groups as the American Management Association (AMA) in efforts to train state department of education officials in the ways and means of introducing MBO into the schools of their states. Such federal policy, of course, is a logical extension of the Ford/Nixon Administration's penchant for cost-accounting and its approach to social problems.

Participants at some of these AMA-state department training sessions were cautioned, "Avoid using such terms as *accountability, MBO,* and *staff assessment* when you talk about this in public or with teacher groups....Call it *planning* and no one will feel threatened."

The most obvious result so far is that teachers get conflicting signals about what all of this will mean to them and to the way they teach. From the higher end of this federal cost-effective continuum, teachers hear the U. S. Commissioner of Education saying, "If properly executed, MBO can be democratic in the best traditions of involving professional colleagues in participatory management." At the other end of this "line" from Washington, in the local school district -- at faculty meetings and via administrative edicts -- teachers are often being *told* what "objectives" they will be held accountable for.

Participatory management, indeed.

---

*A position paper written by Robert C. Snider, professional associate, NEA Instruction and Professional Development.

Increasingly, reports from NEA UniServ directors and local association leaders indicate that this latest brand of USOE-AMA innovation (call it anything but accountability or MBO) is causing as much confusion among district-level school administrators as it is causing among teachers. All too often these administrators view "participatory management" as just a fancy new term for the kind of administration they have always known. Their thinking often goes something like this: "We've always had participatory management in my schools. I manage and my teachers participate. My door is always open."

Among other things, participatory management in schools means that teachers will participate in shaping decisions and in developing objectives they will be expected to use in their work with students. As indicated below, there are many varieties of MBO. But all of them include such common elements as identification of goals, definition of objectives, and some regularly applied product assessment techniques. MBO as a national effort to make schools more accountable may or may not succeed. Certainly, districts should understand the nature and implications of what they are dealing with before attempting to introduce it.

## What Is MBO?

To better understand what happens to the rhetoric of participatory management (which "can be democratic in the best traditions") as it flows from Washington into your classroom, it is essential to know something about the basic ideas behind MBO, where it comes from, where it works, and where it doesn't work.

Management by objectives, as a concept and a term, was first used by Peter Drucker in his book *The Practice of Management* (Harper & Row, 1954), where he wrote:

> What the business enterprise needs is a principle of management that will give full scope to individual strength and responsibility and

> at the same time give common direction of vision and effort, establish teamwork, and harmonize the goals of the individual with the common weal. The only principle that can do this is Management by Objectives and self-control.

It would be difficult for a business leader or an educator -- or anyone, for that matter -- to take issue with such general and high-principled rhetoric about a business enterprise. But Drucker's additive of "self-control" seems to have been lost during the 21 years since MBO was conceived. It may be that, during the interim, "self-control" has been superseded by "participatory management."

Today it is difficult to find specific and detailed definitions of MBO that go beyond the original generalization. As might be expected, the literature of business management contains many examples of this. George L. Morrisey (*Management by Objectives and Results*, Addison-Wesley, 1970) writes that his brand of MBO is "a clear and precise identification of objectives or desired results, the establishment of a realistic program for their achievement, and an evaluation of performance in terms of measured results in attaining them." Morrisey, a management and training executive with Douglas Aircraft, points out that the processes required to implement MBO "are much more complex."

In the world of education, MBO and its meaning are often handled with equally cavalier generalizations. U. S. Commissioner of Education Terrel H. Bell, in his recent book *A Performance Accountability System for School Administrators*, (Parker Publishing Co., 1974), writes about "a variety of significant benefits that result from school management by objectives...a proven systematic approach to educational accountability." Although Bell's book includes a section called "What MBO Will Do" and a chapter on "Getting an MBO System Started," it does not explain in so many words what MBO actually is.

In recent speeches, however, the Commissioner has been somewhat more specific about MBO as he describes USOE efforts to make school management "more resultsoriented." In July 1974, Bell told a meeting of school administrators that "we have been turning more and more to the system of management [MBO] utilized by industry wherein annual goals are set and the entire resources of an organization are focused upon attaining specific objectives."

From the teacher's point of view, MBO in education seems to be another deceptively simple, self-explanatory name for an extremely complex process that has yet to be tried in the schools on a large scale. In these ways MBO is similar to such complex nostrums as individually prescribed instruction and career education. By whatever name, MBO probably will not work in schools without major changes in how schools are organized and in the kinds of decisions teachers can make in relation to what they are expected to accomplish. Certainly, school administrators -- between Commissioner Bell and the classroom teacher -- must be more open with teachers at the outset if a consideration of MBO is to make any sense at all.

The American Management Association plays an important role in spreading the gospel of MBO throughout the business community, and increasingly into schools and other nonbusiness, social agencies. Now in its fourth printing, the AMA best seller *How To Manage by Objectives* (by John W. Humble, 1973) adds nothing to the original 1954 idea:

> At its best, management by objectives is a system that integrates the *company's goals of profit and growth* with the manager's needs to contribute and develop himself personally. *It is not a wonder tool* that can replace intelligent or sensitive leadership, and its misuse can cause *more harm than good*. [Emphasis added.]

Note carefully the last sentence above. It represents the most important lesson learned so far about MBO in the business world. Teachers will be well advised to keep this in mind when their school considers MBO. Although MBO will not replace intelligent and sensitive leadership, it might expose the lack of it. "Accountability," said Commissioner Bell back in 1970, "starts at the top."

All of this leads quite naturally to a closer look at the idea of management, what it means to manage, and who manages what and whom. But management, like MBO, is a concept educators have borrowed from the business community, and the value of this terminology in educational planning will depend on an understanding of the two very different worlds of business and education.

## The Cash Register and the Curriculum

Despite the increased social consciousness of many business leaders, the world which they know best is not the world of education. Therefore, whenever efforts are made to introduce MBO and other "business-tested" techniques into the schools, it will be useful for both educators and business people to keep in mind some of the basic differences between private industry and public education.

1. Ownership. Plants and other property within the business sector are privately owned while most facilities in the educational sector are owned and operated by government in the public interest.

2. Responsibility. Beyond limited government control, a business is responsible only to its directors and to its owners. A public school is responsible to a board of education which represents the general public it serves.

3. Goals. The purposes of individual establishments in each sector are set by their respective governing bodies. For this reason (and for other reasons) the goals and objectives of business tend to be more restricted and more specific than are their counterparts in education.

4. Competition. Some organizations within the business sector compete with each other to attain certain objectives. Organizational survival

and growth, therefore, are usually more important aspects of business goals than is the case in public education.

5. Publics. The size, character, and location of the publics it serves are usually controlled by a business organization, which in some cases may decide to increase the size of its public or market. Schools typically have much less control over such matters.

6. Criticism. Public dissatisfaction with and criticism of business performance is not as widespread, not as articulate, and not as organized as is criticism of schools. Several reasons for this include the fact that the public usually has a much wider freedom of choice among business products and services than it has among schools. Since nearly one of every three people is either a student or a teacher in a public school, schools affect more people directly than does any one industry. And finally, since nearly everyone is either in school or has spent many years in school, the public feels better informed about school matters than it does about making steel, selling insurance, or repairing cars.

7. Products. Products and services are usually much easier to measure, describe, and evaluate in business than in schools.

8. Privateness. Since private enterprise can carry out more of its activities in private than can public schools, it is much easier in business to conceal plans and other managerial activities.

9. Social responsiveness. Due to public pressure and/or legislative action, schools are often required to help solve social problems by setting up new courses of study in such areas as sex education,

driver education, drug education, and racial understanding. Major business involvement in solving such social problems -- with the possible exception of the textbook industry -- is a rarity.

10. Rewards. Employee rewards, incentives, and opportunities for on-the-job growth and rapid advancement are greater and more frequent in business than is usually the case in schools.

11. Research and development. The business sector, taken totally, spends a much larger portion of its budget on research (most of it applied research) and on new product development than does the education sector.

12. Technology. The business sector -- because of its competitive nature, and sometimes with government support -- makes a far greater investment in and use of recent technological innovations than do schools.

13. Advertising. Business is much more concerned with using public relations techniques and the mass media to sell its products and to curry public favor.

14. Change. Business and industry claim to be more attuned to change and more capable of change than are schools. Of course, a major amount of "change" in the business world is concerned only with style, packaging, and other marketing frills.

No value judgments are intended in this brief listing of some of the more obvious differences between business and education. Both types of enterprise continue, in quite different ways, to make important contributions to the public well-being. However, before MBO is transplanted from business and applied as an elixir to improve schools, these basic differences between the two kinds of enterprise should be better understood than is now the case in some school districts.

As might be expected, there are equally basic differences between the role of management in business and the role of management in such nonbusiness social institutions as schools, health care facilities, and law enforcement agencies. Management in business is the name of the game and managers are the most important and most highly paid people in a business organization. Here management has become a science. This is not to say that schools, hospitals, courts, and prisons should not be well managed; the point is simply that social agencies require a kind of management dichotomy, a situation in which professionals such as teachers, physicians, and lawyers must share decision-making responsibility with professional managers (whose sole function is to practice management). Areas of management responsibility in social agencies such as schools, therefore, must be clearly assigned to those individuals who are best equipped to make certain kinds of decisions.

## Managers and Management

All of this raises a fundamental question about MBO in schools: *Is the teacher a manager?* This question is at the heart of much misunderstanding between local teacher associations and school administrator-managers in districts where MBO (by whatever name) is being introduced.

From the classroom teacher's point of view, there are two quite separate areas within an educational system where rather different kinds of management are required. One area is the learning locale -- the classroom -- the point where the goals of the school are transmitted to students. Here the professional teacher clearly has management responsibility. The other area, which might be described as supportive, deals with the many other kinds of management functions necessary for the operation of a complex school system. MBO, at its present stage of development in education, may prove to be more useful at the system level than at the learning level.

*Management* is obviously a key concept in MBO but a relatively new term in educational circles. It is useful to note here how it is most frequently defined in AMA publications and at AMA workshops: "Management is the art of getting a job done through other people." In the growing literature of school management, this function is usually categorized into three kinds of activity: planning, organizing, and controlling. Planning, incidentally, is always cited as the most important element of school management.

With a $55, 68-page three-ring notebook called *Getting Results Through MBO* (1974), the AMA may have established a record of sorts for getting results through MBO. This programmed text draws a fine line between *administration* and *management*. In it we discover that there is now a surplus of administrators and a "crucial shortage of managers." From an organization that derives a substantial income from educational programs to transform administrators into managers, this may make sense. But beyond the administrator-retreading business, there are other indications of an emerging difference between the older idea of administration and the new concept of management.

Teachers who are trying to penetrate the MBO mystique will be interested in this arcane but important difference. According to the AMA, managers actively direct and manipulate the business environment in ways that often involve personal risk. Administrators, on the other hand, passively adapt their work to conform with the environment, usually with little risk, with no rocking of the boat, and with an eye to personal survival. "Managers without personal commitment to risk and the possibility of personal failure are bureaucrats," says George Odiorne, an MBO guru often quoted in AMA literature.

It will come as no surprise that the AMA's use of the term *administration* is not the same as it is in the publications of the older and more pompous American Association of School Administrators (AASA). An AASA publication on MBO favors yet another term -- *executive* (*Management by Objectives and Results: A Guidebook for Today's School Executive,* 1973). Here the AASA explains that "until recently school executives viewed management as a demeaning term that emphasized the mechanical aspects and failed to recognize the leadership dimensions of their positions."

The rare human quality of leadership is as scarce among managers as it is among administrators, although managers -- perhaps because they tend to be younger than administrators -- are more apt to confuse management with leadership. It should be kept in mind that the least remembered American president, Chester Arthur, was the greatest "manager" ever to occupy the White House. He organized the federal civil service system.

With the silly statement about management quoted above, the AASA may have touched inadvertently on an important difference between *management* and *administration*. It has to do with the ways in which, and the basis on which, decisions are made today. Management is not -- or need not be -- more "mechanical" than administration. However, management began as a business term and in many ways it seems more closely related to industrial technology than does the older concept of administration.

On the other hand, management makes the most of information-handling technology in such a way that all useful information is available to decision makers before their decisions must be made. In theory, this is why planning gets so much attention in the management process today. It is now possible to plan with much greater reliability because of a better information base. In business this often makes it possible for managers to anticipate problems and to establish organizational structures that will provide better control over the entire system. The new technology of information processing, storage, rapid retrieval, and instant distribution has in some cases made it possible to delegate management functions more widely throughout a corporate system so that appropriate decisions can be made at the appropriate time and at the appropriate level -- sometimes by a digital computer.

## The Technology of Management

Management, with its information support systems (and such "cookbooks" as MBO), is basically a practical application of systems theory, which in turn is a logical, and probably unavoidable, outgrowth of a technological society. In many ways the modern science of management is dependent on technology. Certainly it is no longer possible to have one without the other -- administrators, executives, and leaders notwithstanding.

This significant fact is usually overlooked in school systems and other more humanistic, nonindustrial social institutions whenever they attempt to borrow and use parts from a very different technological system. To a large degree, the idea of putting MBO into a school system is an example of this phenomenon.

Technology is best understood as a total effort within a tightly controlled (managed) system with goals, objectives, and subobjectives that are unequivocally stated and well understood by all managers. Within this technological milieu such concepts as cost-effectiveness, quality control, MBO, product standardization, and even zero-defects are integral parts of the system. The quality of management at every level in such a system is highly visible. When MBO works well, it works only within this order of systems technology. Since schools are organized and run in quite another way, there is a serious question about the success of MBO in such an educational enterprise.

Present efforts to impose MBO on schools would seem to be an example of putting the cart (of MBO technology) before the horse (of human needs and logic). MBO cannot be usefully applied to schools until schools are reorganized to function as an industrial-technological system. This changes our basic question about MBO to something like: Should massive changes be made throughout the educational system so that it can accommodate MBO effectively? Are such changes possible? Are they necessary? And, will the benefits of MBO be worth the cost?

Such thinking leads to a fundamental question of our time. Can technological models from the industrial sector be applied successfully to social institutions? Technology, simply put, is the organization of knowledge for practical purposes. It generates techniques which increase human capacity to do -- to put a man on the moon, to wipe out polio, to get us to London in three hours, to kill with unbelievable efficiency.

Technology and its management subsystems represent a "*can* do" rather than a "*should* do" approach to human objectives. Since what man can do is not always what man should do, there are several problems related to technology. For example, what do we want to do? What should we do? What will it cost (and will it be worth it)? Who shall decide? Can we anticipate all the results in human terms? Such questions are increasingly important today since it is impossible to develop new technology without a very large national commitment.

Another technological problem that has become obvious during the second half of this century is that, once a new technology is introduced into a society, it tends to become irreversible and pervasive. Technology is amoral and its results, which are usually both positive and negative, are difficult to anticipate and control as they spread throughout society. Guns, drugs, birth control pills -- and perhaps even MBO -- are examples of this phenomenon.

Rather than accepting the results of technology as they are thrust upon us, educational leaders might better serve society by making technology an object of study -- and in the process give technology (and such related processes as MBO) some much needed direction.

To quibble over the terminology of *administration* versus *management* in education today may be needless in those schools where learning is recognized as an objective that is already managed by teachers. However, when management techniques are introduced in an effort to run schools as if they were supermarkets, teachers will want to be well informed about such a development and familiar with its nomenclature.

Words, after all, are what contracts are made of, and words gain meaning and stability only by how they are used, how often they are used, and by whom they are used. By whatever name -- administration, neo-administration, participatory management, cost-effectiveness, or MBO -- a new kind of systematic approach to educational accountability is being borrowed from the world of business and laid on the schools. Although this unholy alliance may "mean business" in a school system, the results thus far at the learning level are experimental at best.

## Some Objections to Objectives

Like the concept of management, the idea of objectives is also central to an understanding of what might be expected from MBO in schools. Certainly it would be difficult to teach anyone anything without first having in mind some kind of learning objectives. And some teachers with specialized training in behavioral science have had a certain amount of success in using behavioral objectives, often in special education. During the past decade, however, a complex "objectives mystique" has grown out of the earlier teaching machine movement, and the situation today is a confusing mixture of behaviorism and humanism. Today any serious consideration of educational objectives can quickly lead into the cultish thicket of behavioral gobbledygook -- "transitional objectives," "turnstile objectives," "behavioral objectives," "functional competence," "principle-centered strategy," and finally that fearsome phrase "terminal objectives."

We often hear that objectives are more singular in business than in education: "Profit is the only objective in business, while schools have a bewildering range of objectives." Teachers who have been asked to write objectives for their students may find small comfort in knowing that the business community also finds the concept of "objectives" to be elusive, confusing, and, according to Drucker, "perhaps even a metaphysical" idea. "It may be as difficult for Management Science to define 'objectives,'" says Drucker, "as it is for biology to define 'life.'...There is no area as yet [in business] where we can really define the objectives, let alone measure the results." (*Technology, Management & Society*, Harper & Row, 1970.)

There are two major difficulties with the use of objectives as control mechanisms for human activity in school or in business. First is the problem of stating and defining objectives. Words are not people, and words are limited in their ability to describe what people do, feel, and think. This limitation is further compounded by the fact that people who write objectives vary in their ability to express ideas in clear English.

Charles J. Hitch, formerly an assistant secretary of defense and a vice-president of the RAND Corporation, has been a leading scientist in the development of systems theory, the grandfather of such things as MBO. Hitch once pointed out that RAND had never undertaken a major systems study where satisfactory objectives could be defined. As a senior member of Robert McNamara's notorious Department of Defense "systems team" during the 1960's, Hitch indicated that uncertainty about goals was to him the "quintessence of military systems analysis." Incidentally, there is now some speculation and study among historians who are trying to determine the role of MBO in our dismal Vietnam involvement.

The second difficulty with objectives is the problem of measurement. An objective is something to be met, to be reached, to be accomplished, to be attained. Therefore, some kind of measurement is obviously needed to determine if and when an objective has been achieved, how well it has been achieved, and who deserves the credit (or blame) for the degree of achievement. Teachers (as well as the NEA Task Force on Testing) will quickly see weaknesses in this approach to accountability in the classroom. Tests used in schools to measure student achievement are increasingly open to question since they may tell us far more about a student's socioeconomic status than about what the student has learned in school. One teacher recently put it this way: "Each of my pupils is a unique reason why MBO won't work here."

*When the ability of teachers to write precisely what it is they want to do for a student is further limited to what can be easily measured, the art of teaching will become less of a profession and there will be no room for such human input as intuition and creativity.*

The "measurement limitation" of objectives (the O in MBO) is well expressed in a four-step fable from the field of systems engineering. It is called "The McNamara Fallacy" and goes something like this: The first step is to measure whatever can be easily measured. The second step is to ignore what can't be measured. Third, assume that what can't be measured easily is not very important. The fourth step is to say that what can't be measured easily really doesn't exist. This, of course, is no way to run a system -- business, military, or school. Should important and complex objectives be abandoned only because they cannot be simply stated and easily measured?

When individual performance is measured and rewarded only on the basis of objectives accomplished, there is a natural tendency (among teachers, students, shop foremen, etc.) to accept only those objectives which can be met with ease. In this climate challenging and difficult objectives are avoided since they will increase one's chance of failure. The result is an organization characterized by trivial objectives, a complacent staff, and mediocre accomplishment.

## The Teacher and MBO

Many teachers have already learned the hard way that MBO in the classroom can mean trouble for everyone concerned. At the outset MBO can be dangerous since it looks so simple and is so easily misunderstood. Like Mary's little lamb, it comes to school as a disarming, simple, and logical idea with great appeal to school boards and to the public. After all, what can be wrong with the straightforward idea of managing by objectives? Isn't that the way most of us manage our own private and public affairs from year to year, from week to week, and from one hour to the next?

In this continual and more-or-less rational process of deciding what we will do before we do it, nearly all of us profit from past experience, and we keep trying to improve our ability to do whatever it is we do. This tendency toward improvement, incidentally, seems to be a universal and genetic characteristic of all living things, particularly humans. We, of course, are different because we can make tools, try new ways of doing things, communicate with each other, and store information. In the process of all this we have produced a Bach, a Shakespeare, a Lincoln, and countless others who have improved the human condition. Obviously humans are more divergent, more unpredictable, and more creative than are trees, bees, or even pigeons.

When MBO becomes immersed in the instructional process as a rigid mixture of behaviorism and business management, it becomes increasingly difficult for teachers to discover and nurture such fragile and unique student gifts as creativity.

Teachers report several major difficulties with MBO as it becomes a more "advanced and sophisticated" part of the instructional process. MBO increasingly stresses conformity and lock-step learning as it discourages deviation (by both teacher and student) from established objectives. Flexibility, measurement, and origin of objectives are all critical factors here, and frequently teachers are not involved in such things.

By the time MBO becomes entrenched as a classroom routine, many of its objectives turn out to be based on cost-effective averages calculated in the business office by "managers of instruction" who are somewhat removed from the arena of learning. Flexibility in this situation becomes little more than a bureaucratic shibboleth involving vast amounts of paperwork. And concerns about measurement soon become the determining criteria for both the origin and the flexibility of objectives.

When this brand of instructional MBO is added to such present school problems as educational neglect, gradedness, and standardized achievement, teachers will lose another degree of freedom to teach -- together with a certain amount of professional self-respect. Because of the often nefarious nature of MBO as a way to manage instruction, the following considerations will be of interest to teachers who have reason for concern about MBO where they teach:

1. MBO is only one aspect of a concerted effort on the part of federal, state, and business agencies to cut educational costs by holding schools accountable for funds spent and for the quality of their products.

2. As a control technique in business and industry, MBO was developed on the "factory" or "beehive" model where workers are managed in an effort to improve their productivity.

3. A major source of difficulty when MBO is transplanted to education arises when school boards and administrators fail to understand the differences between schools and factories in terms of such critical concepts as workers, products, and goals.

4. One result of this misunderstanding of MBO is that attempts are made to diminish the teacher's role as a manager of instruction to the role of a cog in the standardized treadmill of behavioral objectives.

5. MBO tends to remove the instructional decision-making function from the professional teacher, who more than anyone else in a school system has the information required to manage instruction for individual students.

6. Goals, major objectives, and the nature of subsequent educational products should be developed, agreed to, and described in detail at the school system level by the board of education *before* efforts are begun to use MBO at the instructional level. (This, incidentally, is seldom the case.)

7. The system-level goals, objectives, and product descriptions should not be translated into classroom management procedures without the full knowledge and participation of all teachers whose work will be affected by such decisions.

8. Although the ostensible purpose of the categorical control of MBO is to abolish or limit wasteful and unnecessary instructional activ-

ity (i.e., to save money), it can at the same time have a negative effect by limiting tolerance for differences, diversity, and discovery as students (each of whom is a unique human being) strive to understand, to know, and to learn in their own very different ways.

9. An equally limiting effect of MBO can be anticipated in the professional growth of teachers who must work within the narrow intellectual environment of a bureaucracy of objectives.

10. Measurement of the degree to which an objective has been met by each student poses major problems since test results are not good measures of what has been taught -- unless, of course, the test itself has been the course of study.

11. As part of an MBO program teachers are sometimes tyrannized by administrative efforts to evaluate their effectiveness on the basis of their students' test scores.

12. This invalid -- and in some states illegal -- MBO teacher evaluation has occasionally been used as the basis for teacher dismissal, demotion, and other punitive action.

13. A disproportionate amount of responsibility for success is fixed upon an individual teacher who has little control over the variables necessary for such success.

Will MBO work in schools? The father of MBO, Peter Drucker, has often called attention to the impossibility of transferring management techniques, willy-nilly, to nonbusiness institutions. According to Drucker, the first thing a nonbusiness agency must learn from business management is that "management begins with the setting of objectives and that, therefore, noneconomic institutions, such as universities and hospitals, require very different management from that of a business." Drucker, who is occasionally critical of schools ("The dropout is the quality control of education.") takes a dim view of MBO in schools since "there are no measurements for education." (*The Age of Discontinuity*, Harper & Row, 1969.)

A question neglected so far in this paper is: How well does MBO work in the business world? The answer, of course, depends upon who is asked. Independent management consultants tell us that about 70 percent of the business people now attending AMA training sessions on MBO are there to find out why MBO has not worked in their own business establishments. These AMA sessions are not to be confused with the government-sponsored AMA efforts to introduce state education department staffs to MBO.

In the introduction to an AMA publication, *Readings in Educational Management* (1973), John M. Goode states: "The educational administrator who is looking for the latest-model panacea for the ills of education will not find it in the use of improved management practices." MBO is yet another management practice which, for reasons already indicated in this paper, cannot be expected to work very well in schools. Certainly it will not neutralize the effects of bad administrative decisions in a school system.

Can MBO work in the schools? Probably not, says Congressman John Brademas (D.-Ind.), chairman of the House Select Subcommittee on Education:

> ...there are few effective ways of judging the effects of educational expenditures. In fact, we haven't *yet* developed adequate standards of performance for human behavior of most sorts, let alone for the behavior specific to the educational setting. [*CAP Commentary* 1:1; March 1974.]

## The Local Association and MBO

As this paper has indicated, there are many shortcomings to the use of MBO techniques for the management of instruction in schools. This rather obvious truth, however, can easily be lost sight of at school board meetings and in the realities of local politics, where MBO can easily become a lively public issue. In this climate the local teacher association will need to be well informed on the several aspects of MBO so parents and other publics can be told what to expect from this bland, businesslike approach to latter-day behaviorism in their schools.

At present no particular variety of instructional MBO is preferred by those who are attempting to establish it as a means of making teachers more accountable for student productivity. As already noted, such plans can go by many misleading names, and in some cases they are kept "under administrative wraps" as long as possible. Such divisive administration (certainly it isn't good management) does not make for mutual trust, professional respect, or even sound personnel practice. In this environment, association leaders must be alert to spot administrative behavior that will signal a movement toward MBO.

In districts where MBO is just getting started, organized teachers should be aware of such developments. Certainly, teachers should participate and have major influence on decisions that will affect their teaching. This should be accomplished through the process of collective action.

The local teacher association is the most important link between the united teaching profession and the realities of accountability (and its MBO handmaiden) in a school. Although local situations are seldom exactly alike, information about trends, comparable problems, and similar experiences will be helpful to UniServ directors and local leaders. It is important, therefore, that communication channels within the organized profession be used for this purpose.

Of course, the usefulness of accountability information will depend on its relevance to a particular local situation at a particular time. The following suggested positions and program goals for local associations are presented with this in mind. Clearly they will be more appropriate in some districts than in others for establishing strategy and for planning action.

1. The strength of association action is built on facts and detailed information about what is being planned and what has already been done. Rumors and hearsay do not build a good case and can weaken an association's position. Since district-level MBO programs are usually instigated by the state department of education, this agency is a very important -- but not the only -- source of information.

2. In gathering information and data on the extent and nature of MBO in its district, the association will want to concentrate its efforts on the instructional aspects of MBO rather than on such administrative applications as inventory control and bus maintenance.

3. Collective bargaining between school officials and association representatives should be the basis for dealing with any MBO plans.

4. The agenda for meetings which are open to the public should include such questions as:

    a. Is the primary purpose of our MBO effort to improve the quality of instruction or to save money?
    b. No matter how the prior question is answered, what evidence do we have that this can be accomplished by MBO?
    c. Since MBO was developed as an industrial technique with business goals and products, how has (or will) it be adapted for use in our schools?

    d. Have the goals and the major objectives of the school system been agreed upon and clearly defined?
    e. If learning objectives are to be assigned to individual students, how will these objectives be developed and how will their attainment be measured?
    f. How dependable will the results of this measurement be, and how will the evaluative data be used?

5. Regular district level staff planning for MBO should include the early involvement of association-appointed teacher representatives (for various grade and subject areas) who will be active participants when instructional decisions are considered.

6. The ease with which an association can gather information about MBO in its own district will determine the size and scope of its "intelligence gathering" activities. The amount of tension or openness here will be reflected in the tone and nature of the association's public information program.

7. When and if MBO techniques are actually used for major instructional changes, the situation will often represent a giant step backward to behaviorism and the days of teaching machines. In this extreme situation, the association can expect strong support from many parents as well as from groups who understand such problems, who are seriously "anti-Skinnerian," and who want to humanize education.

8. Since most kinds of instructional MBO cannot withstand close scrutiny by experts on tests and measurements, the association will be well advised to emphasize this very weak aspect of MBO, an effort which may require a special study group and outside specialists.

9. If the association does not succeed in efforts to stop MBO before it reaches the classroom, it will then be essential to ease the introduction of MBO by preparing teachers for it. School boards, therefore, should be expected to provide special (in-service) training for all teachers who will be expected to work as a part of the new MBO system with its range of objectives, definitions, and evaluation techniques.

10. As pointed out earlier in this paper, the matter of developing and writing usable "learning objectives" is an extremely technical and specialized professional process, and one for which few teachers have been prepared. Teachers should not engage in this activity without special preparation and/or the continuing services of a qualified specialist.

11. Experts on MBO in business agree that it must have a large degree of built-in flexibility to succeed. The association should see to it that this is also the case with instructional MBO, for flexibility can be a safety valve to protect teachers and students from anticipated (and unanticipated) MBO problems.

12. The association should prepare its members for quick and decisive reaction to badly managed efforts to put MBO in the classroom. In some districts all teachers have been given a list of objectives (objectives they have had no hand in developing) and told to report at once the percentage of their students who will "attain" these objectives by the end of the year. In one building all teachers refused to make such irresponsible guesses. In a nearby high school all teachers did respond -- with predictions of one or two percent.

13. Other associations have countered MBO proposals by "costing-out" in some detail how these new techniques will actually *increase* the total cost of the program.

14. The most dire threat to association members is the tendency on the part of some school administrators to evaluate teachers -- to hold them accountable, and in some cases to dismiss them -- on the basis of their students' test scores in an MBO program effort. The association must be constantly alert to such examples of mismanagement by objectives.

## REFERENCES

American Association of School Administrators. *Management by Objectives and Results: A Guidebook for Today's School Executive*. Arlington, Va.: the Association, 1973. 76 pp.

Bell, Terrel H. "Accountability Begins at the Top." Unpublished speech delivered at the annual meeting of the Education Commission of the States, Miami, Florida, June 21, 1974. 13 pp.

______________. "Management by Objectives: Planning Where To Go and How To Get There." *CAP Commentary* 1: 3; November 1974.

______________. *A Performance Accountability System for School Administrators*. West Nyack, N.Y.: Parker Publishing Co., 1974. 214 pp.

Brademas, John. "Accountability: A Rationale." *CAP Commentary* 1: 1; March 1974.

Churchman, C. West. *The Systems Approach*. New York: Dell Publishing Co., 1968. 243 pp.

Drucker, Peter F. *Management, Tasks, Responsibilities, Practices*. New York: Harper & Row, 1974. 839 pp.

______________. *The Age of Discontinuity: Guidelines to Our Changing Society*. New York: Harper & Row, 1969. 394 pp.

Goode, John M., editor. *Readings in Educational Management*. New York: AMACOM, American Management Association, 1973. 178 pp.

Hartley, Harry J. *Educational Planning-Programming-Budgeting: A Systems Approach*. Englewood Cliffs, N.J.: Prentice-Hall, 1968. 290 pp.

James, H. Thomas. *The New Cult of Efficiency and Education*. Pittsburgh, Pa.: University of Pittsburgh Press, 1969. 65 pp.

Kaufman, Roger A. *Educational Systems Planning*. Englewood Cliffs, N.J.: Prentice-Hall, 1972. 165 pp.

Norman, N. J. *Getting Results Through MBO*. New York: American Management Association Extension Institute, 1974. 68 pp.

## SOME MANIFESTATIONS OF ACCOUNTABILITY
## (And Related Concepts)

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

### Goals - Objectives

- needs assessment
- goal setting
- performance objectives
- outcome specification
- prerequisites
- enabling activities

### Performances

- modules
- technological devices
- behavior modification
- operant conditioning
- mastery learning
- contract learning

### Assessment

- national assessment
- state assessment
- local testing programs
- criterion-referenced tests

### Evaluation

- teacher evaluation
- administrator evaluation
- performance-based teacher education
- program evaluation

### Staff Deployment and Numbers

- differentiated staffing
- team teaching
- numerical staffing adequacy
- teacher-counselor load/class size
- aides and volunteers
- teacher surplus

### Management Systems

- systems analysis
- management by objectives
- program planning evaluation budgeting systems
- performance contracting
- vouchers

### Governance

- accreditation of teacher education programs
- teacher certification and recertification on performance
- tenure
- legal constraints to teacher decision-making authority

# PROBLEMS IN USING PUPIL OUTCOMES
# FOR TEACHER EVALUATION

Robert S. Soar
Foundations of Education
Institute for Development of Human Resources
University of Florida

Ruth M. Soar
Florida Educational Research and Development Council

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

# FOREWORD

Increasingly state association leaders and staff, local teacher groups, and UniServ directors find themselves having to deal with the problems of inappropriate teacher evaluation. As most teacher leaders are aware, one aspect of such problems is reflected in attempts to evaluate teachers on the basis of student achievement.

The attached paper provides a combination of responses to these problems--research findings, technical problems in using test scores, and other considerations. It has been prepared by two nationally eminent researchers in the field. Citing the Soars can serve to increase the credibility of arguments against the use of student achievement for evaluating teachers. Their examples should be particularly useful in dialogues with school district research directors, testing and evaluation coordinators, and other administrators who are committed to using a year's growth in a year as a measure of teacher competence.

In some other NEA material we have called to attention the major reasons why teachers must not be evaluated on the basis of student achievement. Those reasons complement and are supported by the Soar paper and seem worth repeating here in that context:

1. The tests themselves are inadequate for such purposes. Banesh Hoffman has put it well:

> There is no generally satisfactory method of evaluating human abilities and capabilities.... Rough superficial evaluations are of course possible.... But the detection and evaluation of other than superficial ability is inevitably an art demanding insight, taste and knowledge. Current attempts to reduce it to a science and then mechanize it are not only dangerous but in a profound sense unscientific.[1]

---

[1] Hoffman, Banesh. "Psychometric Scientism." Phi Delta Kappan 48: 381; April 1971.

2. The nature of student populations is so varied that outcomes are often more influenced by those variables than by what teachers do. Gene Glass, noted researcher, reminds us:

   > Nothing short of random assignment of pupils to teachers as an iron-clad administrative necessity will ensure that the teachers were in a fair race to produce pupil gains.[2]

3. Many of the conditions which measurably affect learning outcomes are conditions over which teachers have little or no control and they vary widely among schools. Among them are: the number of students teachers must work with each day; time available to teach; planning time; up-to-dateness of curriculum; appropriateness of materials and media; students' physical and emotional readiness for learning; opportunity for teacher in-service education; and most important, decision-making power on curriculum matters.

Each of the reasons cited is considered in one form or another in the Soar paper. And even though some of the technical explanations may go beyond the needs of teacher leaders in responding to the issues, they serve as backups to commonly held teacher association positions.

--Bernard H. McKenna
Professional Associate
NEA Instruction and Professional Development

---

[2] Glass, Gene V. "Statistical and Measurement Problems in Implementing the Stull Act." Mandated Evaluation of Educators: A Conference on California's Stull Act. (Edited by N. L. Gage.) Washington, D. C.: Education Resources Division, Capital Publications, 1973. p. 54.

# PROBLEMS IN USING PUPIL OUTCOMES FOR TEACHER EVALUATION

During the past few years there has been mounting pressure for measuring the outcomes of education, with movement toward holding the teacher, the school, and the school system "accountable" for producing the student learning expected by society. Decreasing enrollments, tighter budgets, and a general trend toward cost effectiveness have added to the pressure.

Measuring pupil achievement has increasingly been proposed as a way of assessing the effectiveness of teaching, and in fact has been mandated by a number of states. This approach is superficially reasonable and attractive, but it is fraught with problems which have not been generally recognized.

H. L. Mencken once commented, "There's always a well-known solution to every human problem--neat, plausible and wrong." The use of pupil achievement as a way of evaluating the teacher, the school, or the school system embodies this misleading simplicity. The solution seems so straightforward: If the job of the teacher is to promote learning in pupils, then it seems reasonable to evaluate the teacher in terms of the amount of learning he produces in his pupils.

The parallel with the industrial setting is clear: If the job of the worker is to assemble relays, then it seems reasonable to count the number of relays the worker assembles and pay him or her accordingly. But in applying this procedure to teaching, a number of problems emerge which have not been widely recognized. The relay assembler receives parts which are identical (at least within very close limits) on which he or she performs a prescribed set of operations, also identical. Then the completed units leave the assembler, again almost identical from one to another.

But none of this is true for the teachers. Pupils appear in the classroom differing in ability, level of achievement, home background, interest, motivation, age--differing in numerous ways. The teacher must recognize these differences as he or she strives to help individual pupils grow toward their own potential. Consequently, the teaching process will differ from pupil to pupil. If the teacher has been successful, each pupil will have improved educationally when he or she leaves the classroom but each will probably be no more like the others than when the year began.

A major dimension, then, of the problem of evaluating teachers in terms of pupil outcomes is the recognition that what goes on in the classroom is not the only, or the most powerful, influence on where a pupil stands in achievement at the end of the year.

## Influences Other Than the Classroom

Research has shown that the differences pupils bring with them when they enter the classroom have significant influence on achievement.

Entry level ability (pretest or fall score) and socioeconomic status are major determiners of what a pupil's standing will be at the end of the school year. These influences probably are more widely accepted than any other, but they are highly interrelated so that one overlaps the other. In practice they cannot be effectively separated.

The fact that IQ and achievement scores in the fall are highly related to spring achievement scores is widely accepted but seldom documented. In a study of 81 fifth-grade classes, Soar and Soar (1973) found correlations between class averages (means) for fall IQ and spring achievement ranging from +.85 to +.90, and correlations between fall achievement and spring achievement ranging from .75 to .85. So the evidence is that as much as 80 percent of the variation in class averages for pupil achievement at the end of the year can be accounted for by pupil characteristics which existed at the beginning of the year, characteristics over which the teacher has no control.

The most extensive data on the influence of socioeconomic status on pupil achievement were presented in the Coleman Report, and more recently and more widely reanalyzed by Mosteller and Moynihan (1972) and Mayeske, et al. (1972). The studies show that as much as 80 percent of the variation in pupil achievement across schools (equal to a correlation of about +.90) can be accounted for by these factors.

Beyond these major influences there are others which help account for differences in pupil achievement and which should be considered. Although the research on family attitudes and support for learning in the home is not as extensive as that for pupil ability (pretest) and social status, it is consistent in indicating relationships between the educational values held by parents and their children's achievement in school. Garber and Ware (1972) found a relation of +.47 between achievement and a combined measure of support for learning in the home for a group of Black and Spanish-American children. All students in the sample met federal poverty guidelines, so that socioeconomic status as usually measured was, in effect, held constant. The same authors cite similar findings from other studies.

Peer group attitude, although again the research is not extensive, has been identified as another important factor which can either support or hinder a pupil's achievement (Anderson, 1970).

Since there is compelling evidence that a number of influences over which the teacher has no control have powerful effects on pupil achievement, it cannot be expected that a teacher will have consistent results with successive groups of pupils. That is, the teacher will not be equally effective in producing growth with all groups because groups differ so widely. Studies by Rosenshine (1970) and Brophy (1972), for example, show that on the average only about 10 to 15 percent of the variation in achievement from group to group reflects the stable influence of the teacher, as shown by a median correlation in the low .30's.

As Medley (1974) has pointed out, and as commonly accepted methods of estimating reliability show (Chronbach, 1960, p. 131), data from about twenty classes would be required for making reliable decisions about individual teachers. Given this requirement necessitating collection of such large amounts of data, using the measurement of pupil achievement as a way to evaluate teachers is impractical as well as invalid.

What these findings seem to indicate is that the education of the pupil is dependent on many conditions in the society, not on the school alone. When the time the pupil spends in the classroom is compared with the time he spends under other influences, and when the degree of influence or control the teacher can exercise is compared with the power of other influences, the limited effect of the teacher is not surprising.

Because influences other than the teacher make a major difference in how much the child learns is not to say that the role of the teacher is unimportant. The teacher is the only formal, institutionalized input the society has to the education of the child and the transmission of an established curriculum. And much of what the teacher does that contributes constructively to the child's future abilities, successes, and satisfactions may not be measured by currently common achievement instruments. It does say, however, that the teacher's influence is limited and that the teacher is most effective when he or she has the support of other elements in the society.

This whole constellation of other influences is usually not given consideration when measures of pupil achievement are proposed as the basis for evaluating teachers. It is reasonable that these influences are strong, since they accumulate over the life of the pupil. It is obvious, then, that pupil standing at the end of any school year is a completely inadequate and even misleading measure of the effectiveness of the teacher or the school. Yet the results of such achievement standings are frequently published by school or by school system.

## "Standing" versus "Change" as Measures of Outcome

"Achievement," which is the most frequently used measure of student learning outcomes, usually refers to the amount of knowledge a pupil possesses at a given point--his or her "standing." The influences cited above show a strong relation to achievement as used in this sense.

An alternative to measuring achievement standing is to measure "change" in achievement from the beginning to the end of the year. When this is done, the influences cited are still likely to have an effect, although to a lesser degree, since change reflects their influence for a shorter period of time.

Although this alternative is appealing as another way of evaluating teaching, it raises still other problems. In a classic volume on the problems of measuring change, Bereiter (1963) commented:

> Although it is commonplace for research to be stymied by some difficulty in experimental methodology, there are really not many instances in the behavioral sciences of promising questions going unresearched because of deficiencies in statistical methodology. Questions dealing with psychological change may well constitute the most important exceptions. It is only in relation to such questions that the writer has ever heard colleagues admit to having abandoned major research objectives solely because the statistical problem seemed to be insurmountable. (p.3)

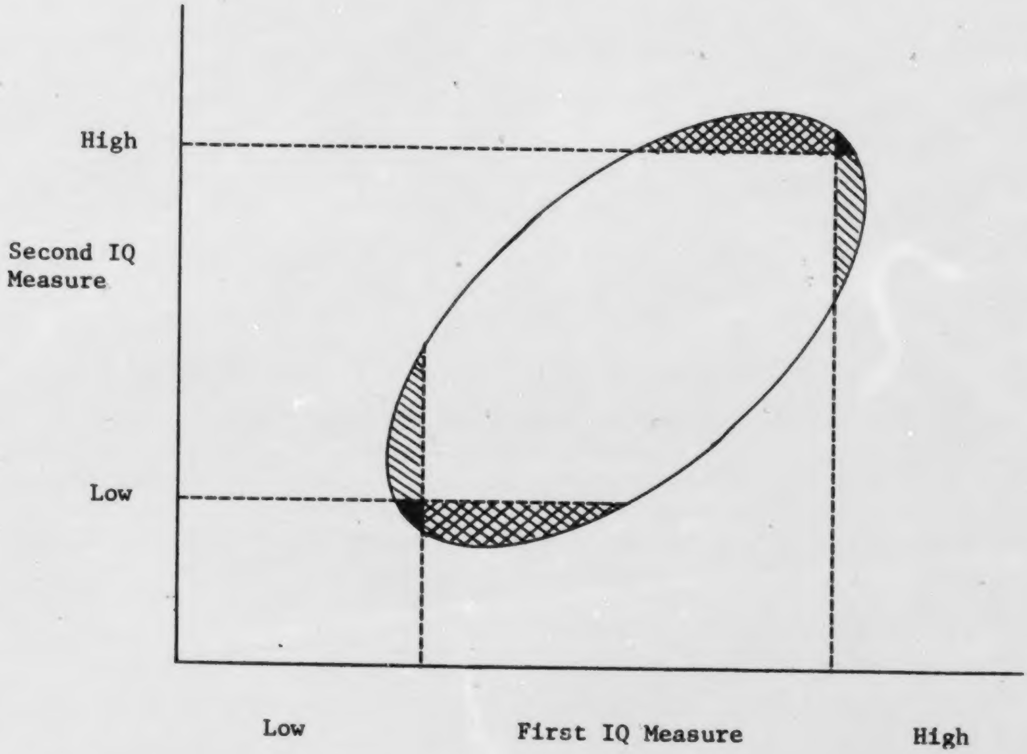
## Difficulties in Measuring Change

If the fall score is simply substracted from the spring score so as to obtain a measure of net change, a new set of subtle but difficult problems is created. An illustration may serve to identify some of them. Figure 1 presents fictitious data from a group of pupils for whom measures of IQ from two forms of a test have been obtained 10 days apart. The initial IQ's are plotted on the baseline and the second IQ's on the vertical axis. Any point in the area outlined by the elipse represents simultaneously the IQ of a pupil on each of the testings, and the high and low 10 percent of the pupils at each of the two times has been indicated by shading and cross-hatching.

It is clear that the pupils who were in an extreme group on the first test were not, for the most part, in an extreme group on the second test. The blackened areas represent the small number of pupils who were extreme on both occasions.

At the upper right, the area is small because the pupils who make the highest scores at any testing are likely to do so on two bases: (1) they are bright (have high verbal skills), and (2) they are "lucky" (that is, they

Figure 1.--An Illustration of Regression Effect


High
Second IQ Measure
Low
Low
First IQ Measure
High

happen to make good guesses on a few items for which they aren't sure of the answer, or the items on this test just happen to be ones for which they know the answers). But they are not likely to be lucky consistently when another form of the test is given, and so on another testing their scores are likely to be lower. Opposite influences will affect pupils at the lower left end of the elipse.

To put it another way, if the cutting point for the top 10 percent is an IQ of 120, there will be a number of pupils with true IQ's close to 120 who will sometimes be above that score on a series of tests and sometimes below it, depending on chance factors. So some fraction of pupils above 120 on the first test will fall below it on the second. Similarly, some fraction of the pupils scoring below 80 on a first test will be above it on a second.

In both cases, extreme pupils have "regressed," or moved, toward the mean. This regression effect can be expected whenever prediction is less than perfect, and the extent of the movement will depend on the inaccuracy of the prediction (Lord, 1963). With most psychological or educational predictions, the regression involved is considerable and may make up a significant proportion of the total range of scores.

The point to be stressed from this example has important consequences: Since pupils who were in the bottom 10 percent the first time were not, for the most part, in that group the second time, they must have moved upward. Similarly, the pupils in the top group must have moved downward. That is, there is a negative relationship between initial standing and the direction in which change is most likely.

As an example of this effect, the pupils who stand highest on an achievement measure at the beginning of the school year will probably show little if any increase in score at the end of the year, and may even show a decline. On the other hand, pupils who score lowest at the beginning of the year will probably show considerable increase. Educators have sometimes been misled by this effect and have assumed that their programs were more functional for low achieving pupils than for high achieving pupils, when in reality all that was involved was the regression effect (the statistical tendency for scores to move toward the average). Similarly, a group of pupils placed in a remedial program because they stand low on a pretest can be expected to show considerable improvement; but again the improvement may be spurious, as a consequence of the regression effect.

This problem creates real difficulties if pupils are tracked on the basis of fall scores and teachers are evaluated on the basis of change in achievement of their pupils. For example, assume that pupils are tested in reading in the fall and the lowest third are put in Miss Jones' class, the middle third in Miss Smith's class, and the highest third in Mrs. Williams' class. We can anticipate that at the end of the year Miss Jones' class will show much improvement and Miss Smith's will show modest gain, but Mrs. Williams will be fortunate if her pupils show any growth at all. The problem is that the gain the pupils show is materially affected by regression effect, so to evaluate the teacher on the basis of pupil gain would be manifestly unfair.

There are statistical procedures for attempting to eliminate this effect, but as Bereiter (1963) commented, it is impossible to be certain that appropriate adjustments have been made; and the expertise to do even the best that can be done with the problem is not widespread. And, of course, all the out-of-school influences on achievement standing discussed earlier also influence gain, although to a lesser degree. So it is clearly inappropriate to use pupil change as a way of evaluating teachers where a teacher may suffer as a consequence of the error involved.

Teacher Performance Tests. A procedure for evaluating teachers which attempts to bypass the problems of change is the performance test or the evaluative teaching unit (Flanders, 1974). In it, the teacher teaches a prescribed brief unit (sometimes as little as a few minutes or as much as two weeks) and pupil knowledge is then tested. The attempt is made to minimize the problems of measuring gain by teaching material in which pupils should have little or no preknowledge, so that all presumably start at "ground zero." But the other problems of using pupil achievement to evaluate teachers still apply. In addition, there are questions of whether teaching material which does not have to be integrated into previous knowledge requires the same skills as the usual teaching setting and whether such short-term learning generalizes to long-term learning. There is the final difficulty that the performance of teachers on a unit of a few minutes does not predict their performance on a two-week unit (McDonald, 1974). Assuming that either can be used to predict year-long performance then seems risky.

Even if the measurement of standing or gain in achievement were a satisfactory way of evaluating teachers, there is still the problem of selecting the objectives to be measured.

## What Objectives Should Be Measured?

Although subject matter achievement has been the primary focus of the discussion thus far, it is clear that schools are charged with and have accepted some degree of responsibility for many other kinds of pupil growth.

The Need for Multiple Measures. Over a long period schools have given attention to the social development and the moral values of pupils. And a broad view of the relationship between school and society suggests that when a problem emerges in the society, one of the first steps is likely to be to involve the school in solving the problem. Traffic problems led to driver education; a concern for the loyalty of government employees led first to a ban on teaching about communism in the schools and later to the requirement that it be taught; problems of drug abuse have led to drug abuse education in the schools; concern about sexual attitudes has led to sex education; concern for occupational choice has led to career education in the schools; and when concern for segregation of the races became pressing for the society, the first and the major attempt to deal with the problem was delegated to the schools. To evaluate teachers and schools solely on the basis of the subject matter gains made by pupils grossly under-represents the broad range of objectives for which teachers and schools have been given some degree of responsibility. Yet for many of these objectives there are no measures which are immediately, for some even remotely, available.

Simple Versus Complex Learning. Even within the subject matter realm there are problems which are largely ignored. One of these problems is the need to distinguish complex achievement growth from simple growth and to provide appropriate measurement for each. Memory of facts (rote memory) falls at the simplest level and complex problem solving, abstracting, and generalizing fall at the complex level. The distinction is between retrieving information (memory) and processing information in its varying degrees of complexity. There is some evidence from a number of studies that the teaching behaviors which are associated with greatest growth in simple tasks are different from those which are associated with greatest growth in complex tasks (Solomon, Bezdek, and Rosenberg, 1963; Soar, 1968; Soar and Soar, 1972, 1973).

Most studies of pupil achievement fail to make this distinction; and the current stress on criterion-referenced measurement, emphasizing "small-step" learning, seems likely to focus on simple kinds of learning. Measures of complex learning are slow and difficult to construct, in contrast to measures of simple learning, which can be more easily and quickly developed. Evaluating all subject matter at all grade levels would almost certainly require the construction of many new measures which would likely emphasize simple kinds of achievement, given the ease with which they can be constructed and the emphasis on criterion-referenced measurement. If teachers were to be evaluated on the basis of pupil achievement, then, it seems likely that the teacher who emphasizes simple learning would be more positively evaluated than the teacher who emphasizes more complex learning. This would be an unfortunate result.

A further problem related to the difficulty of measuring complex achievement growth is the likelihood that some highly valued objectives grow too slowly to show change within a school year -- objectives such as complex problem-solving skills, citizenship, attitudes, learning to get along well with others, and creative expression. On the other hand, it seems likely that measures of short-term learning would tend to emphasize simpler kinds of learning.

## Other Problems in the Use of Pupil Outcomes

A description of an application of accountability in England a century ago makes one of the problems clear (Small, 1972). In that setting, teachers were evaluated on the number of their pupils who attained the minimum level of achievement expected for the particular grade. The result was that teachers concentrated their efforts at the minimum level of proficiency, with a consequent lowering of the quality of instruction.

Another problem of serious consequence in the use of pupil measures is raised by the OEO study of performance contracting, which found that the superior achievement of performance contracting programs disappeared when the teaching was controlled to eliminate the possibility of teaching the test (Page, 1972, 1973). The implication seems clear that, in a setting in which financial return follows from pupil achievement, teaching the test is likely to occur at least a portion of the time. This is a very reasonable finding and one which is well known, even in cases where a financial return is not involved -- teaching to the Regents Examination, for example.

A final problem is the possibility of bias if the teacher is the test administrator. Even outside test administrators have difficulty not "helping" pupils; but where a teacher is affected personally, it seems possible that his or her behavior might be influenced, even though unconsciously. This problem could be dealt with by using only specially trained test administrators, but this could be very costly.

## Summary

When all these problems in the use of pupil achievement for teacher evaluation are considered, they become overwhelming. The influence of the teacher is minor compared to out-of-the-classroom influences -- pupil ability, previous knowledge, the home, the peer group, motivation, and others. What the pupil brings to the classroom in this respect is clearly a much stronger determinant of where he or she will stand at the end of the year than anything that has been done in the classroom. Influences on the development of future achievement measures seem likely to limit them to relatively simple measures for some time to come. Tests available for measuring the other objectives for which the teacher is to some degree responsible are relatively few. In addition to these problems, there are statistical difficulties in the measurement of change which are extremely serious, if not disabling. They are still further exacerbated by the likely problems of teaching the test, of the teacher giving attention primarily to a small portion of the students, and of obtaining valid measurement in the classroom.

Taken all in all, this is an imposing array of difficulties, most of which have gone unrecognized when it is proposed that teachers be evaluated by measuring the outcomes of their pupils.

# REFERENCES

Anderson, G. J. "Effects of Classroom Social Climate on Individual Learning." American Educational Research Journal 7: 135-53; March 1970.

Bereiter, C. "Some Persisting Dilemmas in the Measurement of Change." Problems in Measuring Change. (Edited by C. W. Harris.) Madison: University of Wisconsin Press, 1963.

Brophy, J. E. Stability in Teacher Effectiveness. R & D Report Series 77. Austin: Research and Development Center for Teacher Education, University of Texas, July 1972.

Cronbach, L. J. Essentials of Psychological Testing. Second edition. New York: Harper and Brothers, 1960.

Flanders, N. A. "The Changing Base of Performance-Based Teaching." Phi Delta Kappan 55: 312-15; 55: January 1974.

Garber, M., and Ware, W. B. "The Home Environment as a Predictor of School Achievement." Theory Into Practice 11: 190-95; June 1972.

Lord, F. M. "Elementary Models for Measuring Change." Problems in Measuring Change. (Edited by C. W. Harris.) Madison: University of Wisconsin Press, 1963. Chapter 2, pp. 21-38.

McDonald, F. J. "The State of the Art in Performance Assessment of Teaching Competence." Performance Education: Assessment. (Edited by T. E. Andrews.) Albany: Multi-State Consortium on Performance-Based Teacher Education, New York State Education Department, 1974.

Mayeske, G. W., and others, A Study of Our Nation's Schools. U. S. Department of Health, Education, and Welfare, Office of Education, Report No. DHEW-OE-72-142. Washington, D.C.: Government Printing Office, 1972.

Medley, D. M. "Research and Assessment in PBTE." AACTE Leadership Training Conference on Performance-Based Teacher Education. St. Louis, April 30, 1974.

Mosteller, F., and Moynihan, D. P. On Equality of Educational Opportunity. New York: Random House, 1972.

Page, E. B. "A Final Footnote on PC and OEO." Phi Delta Kappan 54: 575; April 1973.

__________. "How We All Failed at Performance Contracting." Phi Delta Kappan, 54: 115-17; October 1972.

Rosenshine, Barak. "The Stability of Teacher Effects Upon Student Achievement." Review of Educational Research 40: 647-62; December 1970.

Small, Alan A. "Accountability in Victorian England." Phi Delta Kappan 53: 438-39; March 1972.

Soar, R. S. and Soar, R. M. *Classroom Behavior, Pupil Characteristics, and Pupil Growth for the School Year and for the Summer*. Grant numbers 5 ROI MH 15891 and 5 ROI MH 15626, National Institute of Mental Health, U. S. Department of Health, Education, and Welfare. Gainesville: University of Florida, 1973.

________________________. "An Empirical Analysis of Selected Follow Through Programs: An Example of a Process Approach to Evaluation." *Early Childhood Education*. (Edited by I.J. Gordon.) Seventy-first Yearbook, Part II, National Society for the Study of Education. Chicago: University of Chicago Press, 1972. Chapter 11, pp. 229-59.

Soar, R. S. "Optimum Teacher-Pupil Interaction for Pupil Growth." *Educational Leadership Research Supplement* 2: 275-80; December, 1968.

Solomon, D.; Bezdek, W. E.; and Rosenberg, L. *Teaching Styles and Learning*. Chicago: Center for the Study of Liberal Education of Adults, 1963.

# ROLES AND RESPONSIBILITIES OF GROUPS CONCERNED WITH STUDENT EVALUATION SYSTEMS

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

# PREFACE

The testing movement in this country seems nearly to have gotten out of hand. To contribute to bringing under control this rampaging enterprise, the NEA Task Force on Testing* identified groups concerned with the evaluation of student progress who needed to assume some responsibility for assuring that such evaluation is constructive, manageable, and useful. The Task Force gave the major leadership and input to the development of the document which follows.

Unquestionably, were all the groups cited to take seriously and act in accordance with the roles and responsibilities here attributed to them, many of the present problems surrounding the testing movement would be reduced significantly.

Bernard H. McKenna
Professional Associate
NEA Instruction and Professional Development

---

*The Task Force on Testing consisted of eight NEA-appointed practitioners who studied the problems of testing over a period of three years. Their findings and recommendations are available in three reports to the Representative Assembly (1973, 1974, and 1975).

# ROLES AND RESPONSIBILITIES OF GROUPS CONCERNED WITH STUDENT EVALUATION SYSTEMS*

The roles and responsibilities delineated below for specific groups of persons particularly concerned with student evaluation are based on findings of and positions taken by the NEA Task Force on Testing. These recommended roles and responsibilities are considered essentials for achieving the goals of --

- Sound and fair development of evaluation systems;
- Appropriate distribution and administration of evaluation systems;
- Accurate and fair interpretation of the results; and
- Relevant and constructive action programs based on the results.

## A. TEACHERS, INDIVIDUALLY OR COLLECTIVELY THROUGH THEIR ASSOCIATIONS, SHOULD, AS APPROPRIATE:

1. Seek representation on school district, testing industry, and government (state and federal) decision-making groups for test development (e.g., Educational Testing Service, National Institute of Education), become involved in item analysis and selection, and provide feedback on content and format of tests.

---

*The term evaluation systems is used instead of tests because it is believed that a wide variety of alternatives to tests should and can be developed through research and tryout leading to their validation for evaluation purposes.

2. Plan and negotiate for, or otherwise reach agreement with the school administration on, released time and district in-service education programs to prepare members in the use of tests.

3. Plan professional activities in the area of testing for all members of the association.

4. Seek and participate in in-service training in the area of testing to learn to construct and evaluate teacher-made tests, to learn about objective- or criterion-referencing, to learn about alternative assessment tools, to learn appropriate reporting procedures, to develop an awareness of the variety of tests and their purposes, to keep abreast of latest research findings, and to develop the ability to analyze and criticize standardized tests as they relate to school and district programs and goals.

5. Work to influence test makers and the local and state school systems and secure from them a firm commitment to evaluation programs that will lead to the improvement of instruction.

6. Keep parents and other interested community groups informed about trends and promising developments in evaluation procedures and about unsound testing practices.

7. Negotiate for, or otherwise reach agreement with the school administration on, provisions guaranteeing teacher lead time for preparation for testing, appropriate testing conditions and scheduling, and follow-up time for scoring. Provisions should spell out teachers' appropriate role in the test-scoring process (e.g., to remedy the inordinate amount of time spent on hand scoring).

8. Thoroughly familiarize themselves with tests to be given (assuming they have been furnished with appropriate background materials and sufficient time for learning about administration of the instruments).

9. Develop an understanding of their students' cultural and socioeconomic backgrounds and sensitivity to their individual needs and problems in order to avoid the possibility of irrelevant and biased testing.

10. Periodically review tests to determine their relevancy to instructional goals and objectives and their timeliness, and recommend to the school administration and the testing industry abandonment of irrelevant and outmoded tests.

11. Secure by appropriate means -- from the school or school district administration, as deemed necessary -- the right to determine what tests will be administered, when they will be given, and at what intervals. They should also secure the right to determine exemptions from testing.

12. Secure by appropriate means -- from the school or school district administration, as deemed necessary -- the right to determine proper physical arrangements and time frames for testing as appropriate for themselves and their students. Time allowed should be sufficient for thorough orientation of students to the test being given, and for scoring and reporting results.

13. Be responsible for providing a nonthreatening attitudinal atmosphere for students during testing sessions, given the proper conditions.

14. Assure that machine-scored results are validated by hand scoring a sample of tests.

15. Take an objective approach in interpreting test results, never using them as a weapon against students.

16. Seek to ensure that test results are not used to categorize students into homogeneous groups or as a criterion for student admission to programs of their choice.

17. Strive for accuracy in interpreting test results, relating them to socioeconomic factors affecting individual students.

18. Have respect for student privacy in interpreting test results and manifest that respect by working to secure school district policies guaranteeing students' privacy in the reporting and dissemination of test results, which should not be for public information.

19. Urge strict enforcement of the federal Privacy Act affecting pupil records.

20. Work to secure legislation which will prevent publication of test scores.

21. Work to secure legislation which will prevent the use of test results as a basis for allocation of local, state, or federal educational funding.

22. Assure that test results are not compared among classrooms or buildings or with other districts or regions.

23. Report on test results in a manner appropriate to a varied audience -- students, parents, media, professionals.

24. Recommend general and specific program improvements to the school and school district administrations, and to effect the improvements, identify the needed resources and remedial measures and programs.

25. Secure through the appropriate means -- from the school or school district administration, as deemed necessary -- the stipulation that test results will not be used in evaluating teacher performance. (Teachers should be held accountable for conducting the best instructional process possible under existing conditions, not for guaranteeing learning.)

26. Take a position in favor of the inclusion of courses in tests and measurements in all teacher preparation programs, and provide input on testing problems and issues to their representatives on professional governance boards or commissions to help in the formulation of standards and requirements for teacher education and licensure. (There is little evidence that most preparing institutions or states specifically require or encourage classroom teachers to acquire the knowledge and skills necessary for using tests.)

B. OTHER PROFESSIONAL ASSOCIATIONS SHOULD:

1. Search out and synthesize information on all issues associated with the development, use, and abuse of tests and communicate to the members any information affecting them or their students.
2. Organize study committees of members knowledgeable in testing to develop policies, guidelines, and procedures for testing. Such committees should seek input from all members and consultation from experts in the field.
3. Serve in a "watchdog" capacity on the introduction and administration of curriculum-related tests to assure their appropriateness for schools, and communicate regional concerns to the testing industry.
4. Pursue needed changes in school curriculum programs as identified through the results of testing, this in cooperation with other associations in the region which represent comparable educational and socioeconomic conditions.
5. Identify alternatives to standardized testing.
6. Provide background information and regional concerns to those responsible for drafting or introducing state legislation, and work for passage of legislation to regulate types of tests and uses of the results. These efforts should include calling for the testing of students in their dominant language (except, for example, proficiency tests in English).
7. Urge strict enforcement of the federal Privacy Act affecting pupil records.

C. STUDENTS, INDIVIDUALLY OR COLLECTIVELY AS APPROPRIATE, SHOULD:

1. Seek a role in the development of tests through representation on school, district, and testing industry committees and by providing feedback on test content and format.
2. Take positions against the use of measurement instruments that they feel are biased and will lead to unfair results on the basis of race, sex, socioeconomic status, language, or culture, and make these positions known to the school and school district administration and the testing industry.

3. Make every effort, assuming they have been afforded proper orientation, to thoroughly understand the purpose, and intended uses of results, of any test to be administered in which they will be involved. Students should have the right to refuse to take a test which is known to be racially, culturally, or otherwise biased.

4. Seek a role in determining the conditions of test administration -- including scheduling, preparation, length, location, facilities. (Many tests are administered under adverse conditions, with little attention given to the total physical environment and insufficient time allowed for orientation.)

5. Call attention to any physical or attitudinal pressures in the administration of tests which they feel threaten them or their performance.

6. Insist that they be given a thorough explanation of test results in a meaningful way and in language they can understand.

7. Take a position on the use of test results, demanding guarantees of privacy and the right to determine to whom the results will be released, insisting that results not be used to demean or categorize them or to deny them admission to programs of their choice, and urging strict enforcement of the federal Privacy Act which affects pupil records.

8. Seek a role in deciding on alternatives for meeting student needs as identified through the results of testing, insist on the right to choose from among alternatives, and become involved in the planning of remedial programs.

Local Student Action for Education and Student NEA groups might assume the leadership role in involving all students in the evaluation programs of the school or school district and serve as the voice of student opinion and the vehicle for their protection against the adverse effects of evaluation.

D. MINORITY GROUPS SHOULD:

1. Actively seek representative involvement on testing industry and school system decision-making groups for test development and use.

2. Urge test makers to (a) revise tests in consideration of minority differences, eliminating culture-related items from current tests and working toward cross-cultural instruments; (b) research ethnic and regional test requirements and withdraw tests found to be inappropriate to the population being tested; and (c) explore and recommend alternative forms of student evaluation.

3. Request from the testing industry documentation on norming procedures and population bases for norming.

4. Keep members informed of improper test procedures and seek support or legal assistance where tests and results are misused.

5. urge minority students to refuse to take tests which are found to be biased and urge minority teachers to refuse to administer such tests.

6. Work to prevent the invasion of student privacy in interpretation and use of test results.

7. Work for legislation to prevent publication of test scores and for enforcement of the federal Privacy Act affecting pupil records.

8. Promote legislation to prevent the use of test scores as a basis for allocation of local, state, or federal educational funding.

9. Take strong positions and action against the use of test results for tracking, to denigrate minority intelligence, or to deny students entrance to programs.

10. Expose the erroneous contentions of Shockley and Jensen that some groups in society are genetically less intelligent that others. (The typical group test is considered (a) an unreliable measure of mental ability and (b) to be biased against minorities, having been standardized on a different kind of population.)

11. Actively seek changes in curriculum (including textbooks) to reflect minority concerns and diagnostic services based on student needs as identified by appropriate testing.

12. Seek community support and funds for appropriate new or experimental education programs based on needs identified through means other than testing.

13. Become involved in planning and providing pre- and in-service education for teachers to orient them to minority problems and needs related to testing.

14. Seek public awareness of and concern for minority problems in testing, and pressure community media to help keep the public informed, especially on issues related to proper interpretation and use of test results.

15. Form coalitions for action in the development and use of tests.

E. THE TESTING INDUSTRY SHOULD:

1. Include in test development substantial numbers of persons from all groups that have an interest in and knowledge about testing, particularly representatives of classroom teachers and minority groups.

2. Be responsible for producing culturally fair and bias-free tests that contain relevant items.

3. Work with all concerned groups in constantly monitoring, updating, and revising their tests. The industry should immediately withdraw out-of-date tests from the market, as recommended by those who use them.

4. Take regional diversities into consideration in constructing tests to ensure relevance of test items.

5. Correlate tests to current and developing curricula.

6. Improve sampling techniques and broaden sampling bases.

7. Undertake in-depth research and development to perfect a wide variety of alternatives to standardized norm-referenced tests.

8. Provide with each test copy a cover document specifying what the test is designed for (to reveal depth of subject knowledge, to verify reading comprehension, to establish equivalency, etc.) and what groups (e.g., "early childhood," "later elementary") it is appropriate for. The document should also include a release form for student signature testifying that "I understand the purpose of the test..." or "I am taking test under protest...."

9. Provide an up-to-date manual with each standardized test, issued in English and other appropriate language editions depending on the student population. The manual should give clear and complete information for administration of the test, including proper physical arrangements; define proper and improper uses of the test, warning particularly against using the test for purposes of teacher evaluation; explain various ways of interpreting results, providing information on the basis of norming to ensure proper interpretation and including a "Surgeon General's warning" on the dangers of misinterpretation; delineate limitations of the test.

10. Provide with each test, not just bench marks, but a range of scoring norms.

11. Constantly monitor the distribution of standardized tests to ensure proper use, respond promptly to charges of misuse, and refuse to sell tests or report scores where misuse is evident.

12. Provide in-service training for teachers and administrators in the use of standardized tests; provide consultants and test administrators to assist teachers in giving tests and developing sensitivity to testing conditions; and have representatives available as resource persons for interpretation of test results.

13. Provide information on the use of standardized tests and interpretation of results to schools of education and urge them to include courses in tests and measurements in their required professional preparation for teachers. Such courses should include instruction on limitations of tests, potential bias, and a broad range of alternatives to testing.

14. Develop recommendations for curriculum revisions as related to test results in order to help teachers in planning remedial programs for students.

15. Establish an extensive PR program to keep the public informed on testing issues and developments, issuing information materials in English and other language editions.

F. SCHOOL ADMINISTRATORS SHOULD:

1. Ensure that, when appropriate, all tests to be administered reflect the uniqueness of the geographic region in which they are administered and that locally developed and standardized tests reflect updated curriculum.

2. Involve teachers, students, and parents in decision making related to the testing program.

3. Ensure that all teachers who must administer tests are provided with adequate supplies for the students, proper physical arrangements, and thorough orientation time, including practice testing.

4. Provide released time for teachers for in-service training in the administration of tests.

5. Ensure that test results are not used to label students, that the confidentiality of scores is protected in a professional manner, and that the federal Privacy Act affecting pupil records is enforced in school buildings and districts.

6. Make available to teachers or specialists tools for diagnostic purposes and training in their use.

7. Keep parents informed about test results (using nontechnical language) and keep the school board informed about the limitations and possible misuses of tests.

8. Continually evaluate the total testing program.

G. APPROPRIATE COLLEGE AND UNIVERSITY PERSONNEL SHOULD:

1. Serve a research function, providing to NEA and other concerned groups and to faculty in the school of education their findings on the use and misuses of standardized tests (including their own testing devices), test bias, and alternatives.

2. Serve in a consultative capacity to the testing industry, providing information on student population and needs, new curricula, college admission policies, scholarships, equal opportunity programs, and the like.

3. Serve in a consultative capacity to school systems for in-service teacher education and for decision making about curriculum changes based on the results of testing.

4. Seek the involvement of practitioners in decision making relating to professional preparation in tests and measurements.

5. Monitor test results from school districts in their region in relation to new directions for open admissions, equal opportunity programs, scholarships, etc., and keep junior and senior high schools informed about the relationship of test scores to admission policies and program choice.

6. Form coalitions to influence legislation and provide expert testimony on the proper uses of tests and test results.

## H. GOVERNMENT AGENCIES

1. The U. S. Congress should legislate restraints on the use of tests that prevent equal educational opportunity.

2. The appropriate federal agencies should:

   - Provide quality control of testing by taking steps to restrain the testing industry from publishing tests that are improperly constructed and by monitoring instruments to ensure their constant updating.

   - Provide technical assistance and information to educators and the public regarding test development and use.

   - Increase research efforts in standardized tests and alternatives.

   - Assure that teachers are involved in decision making about the use of revenue-sharing funds as they apply to the school system's testing program.

3. State education agencies should:

   - Provide consultant services, financial assistance, and models for quality in-service education for teachers on the proper administration of tests and on limitations of test results.

   - Provide for alternatives to standardized tests for state assessment programs.

   - Prevent the improper distribution and administration of large-scale assessment program materials by instituting sampling procedures as opposed to blanket testing.

4. Local education agencies should:

   - Provide released time and quality in-service education for teachers and other school personnel on the administration of tests and use of results.

   - Prevent misuse of large-scale assessment instruments by instituting sampling procedures as opposed to blanket testing.

5. Education agencies at all levels should:

   - Involve teachers in decision-making on test development.

   - Provide the funds for innovative programs to develop alternatives to standardized testing and interpretation.

   - Provide the funds for long-range experimental testing programs.

# WHY SHOULD ALL THOSE STUDENTS TAKE ALL THOSE TESTS?

## (EVERY-STUDENT TESTING OR SAMPLING OF SELECTED GROUPS?)

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

# WHY SHOULD ALL THOSE STUDENTS TAKE ALL THOSE TESTS?

## (EVERY-STUDENT TESTING OR SAMPLING OF SELECTED GROUPS?)

The NEA Task Force on Testing, in its first interim report, states:

> The Task Force believes there is overkill in the use of standardized tests and that the intended purposes of testing can be accomplished through less use of standardized tests, through sampling techniques where tests are used, and through a variety of alternatives to tests....
>
> Representatives of the testing industry and others told the Task Force that sampling of student populations could be as effective as the blanket application of tests that is now so common. Some suggested that such procedures, in addition to increasing the assurance of privacy rights, would conserve time, effort, and financial expenditure.[1]

The blanket use of tests (every-pupil testing) in some state assessment and local testing programs appears to require inordinate amounts of time and resources on the part of teachers, other personnel involved in test administration and interpretation, and the students themselves.

Criticisms of the blanket use of tests have come from a variety of prominent researchers, evaluators, and other educators.

House, Rivers, and Stufflebeam, in their evaluation of the Michigan accountability system, concurred that in that state:

> Statewide testing as presently executed also raises the question of the feasibility of every pupil testing. This practice appears to be of dubious value when the cost of such an undertaking is compared with the resulting benefits

---

[1] In Task Force and Other Reports presented to the Fifty-Second Representative Assembly of the National Education Association, July 3-6, 1973, Portland, Oregon. pp. 26-46.

> to local level personnel.... The local, and hence overall, costs could be reduced by a matrix sampling plan which requires that each student tested take only a few items.... In the long run, a matrix sampling plan will be the only one feasible from a cost and time standpoint. The cost and time required for every pupil testing for the whole state would be horrendous.... We feel that it /strict adherence to a statewide testing model/ will result in useless expenditures of monies and manpower, in addition to producing unwarranted disruptions of the educational programs within a great number of schools.[2]

In a paper entitled "Criteria for Evaluating State Education Accountability Systems," the National Education Association has laid down fifteen basic principles, one of which is as follows:

> If the state desires test data for its own planning purposes, it should use proven matrix sampling techniques which will not reveal schools and which will greatly reduce costs.
>
> Matrix sampling techniques can give an accurate picture of the state by various categories much more efficiently than testing each child with an entire instrument.[3]

It was with such admonitions as these in mind that this paper was developed. And while some procedures are appropriate for evaluating all students in one way or another for particular purposes, it would appear that there is gross over-use of blanket testing procedures.

To help teachers and other educators better understand some main considerations related to sampling, the NEA obtained permission from Dr. Frank Womer, Michigan School Testing Service, University of Michigan, to reproduce material from a monograph of his on developing assessment programs.[4] In addition, Dr. Womer prepared, especially for this paper, a section on item sampling. Dr. Womer's recommendations follow.

---

[2] House, Ernest R.; Rivers, Wendell; and Stufflebeam, Dan. An Assessment of the Michigan Accountability System. Michigan Education Association and National Education Association, March 1974. pp. 14-16.

[3] National Education Association. "Criteria for Evaluating State Education Accountability Systems." Washington, D. C.: the Association, n.d...

* * * * * * * * * * * * * * * * * * * * * * * * * * * *

## Determining Whether Sampling Is To Be Used

The decision whether to test an entire population or use a sample involves a combination of concerns. Clearly there are policy considerations; clearly there are psychometric[5] considerations; clearly there are data collection considerations; and clearly there are cost considerations. The best possible staff and consultant thinking on this question should be brought to an advisory committee for them to consider very carefully.

Probably the most crucial consideration is a policy one, since psychometrics, data collection, and cost generally would argue on the side of sampling rather than using an entire population. If it is deemed wise for policy reasons to test all students in a population, that preference, typically, will have to be weighed against available resources and technology; so we will consider first the policy implications of the two choices.

One needs to look carefully at the purposes and goals of a specific assessment program in determining whether sampling is appropriate. If all of the specific purposes and objectives of an assessment program can be met by group results, then sampling must be considered.

---

[4] Womer, Frank B. Developing a Large-Scale Assessment Program. Denver: Cooperative Accountability Project, 1973.

[5] Editor's note: Psychometrics in the strictest sense of the definition has to do with the measurement of mental abilities. It has come to be used much more broadly to define a wide range of activities in assessment and evaluation.

The only assessment situation that clearly calls for common data collection on all members of the population is when it is deemed essential, for improved decision making, to have exactly the same test information for every pupil in a given grade in a state (or other assessment unit). It is exactly this situation that has prevailed for years in local school districts that have every-pupil achievement or ability testing at some grade level. Historically, the compulsory state testing programs were examples of this situation; the voluntary programs were not. If a state mandates common testing for all students it is taking over a role that local districts traditionally have held. This may be good or this may be bad depending on one's point of view of the role of a state department of education. It certainly has important policy implications.

There are many facets to this point, but it should be kept clearly in mind that it is not necessary to test every pupil at a given grade level on identical material in order to get a good picture of education outcomes of groups of students; it is necessary only if one feels that each teacher in an entire state at a given grade level must have the same information for each pupil.

Probably the greatest advantage of sampling is that for a given amount of effort (and money) one can gather more usable information than by using an entire population. If the goals of an assessment program are to gather statewide information only, it is hard to conceive of any reason for testing all students in a given grade. For example, if there are 50,000 third-graders in the state of Limbo, and one wants to gather state statistics only, it is very possible that a sample 5,000 students (or even 500) would be sufficient _if_ they are selected by a probability sample....[6] Or, if one can afford to test all 50,000 third-graders, and if it is deemed wise to do so, one could select ten 5,000-pupil samples and secure information on ten subject areas, or one could go into great depth of information gathering in two or three subject areas. The combinations of possibilities of sampling pupils and content are almost endless.

If one wants district-level information, then sampling becomes a different situation. In a school district with one third grade, sampling of pupils is hardly possible for most assessment purposes. In school districts with many third-graders, sampling could provide a greater variety of information than common testing on every pupil, in the same fashion as at the state level. Specific decisions of how far to carry sampling should be made only after advice from a sampling statistician. Sampling is a highly developed technical field,and the implications of any decisions to sample or not to sample must be reviewed by competent samplers.

Other "compromise" possibilities exist. One could test all students in a population with one short test, while using a sampling approach for other tests. This approach would provide some common information on all students but would allow for greater depth of data collection over a subject area.

> Principle: Sampling of pupils and/or content should be given very serious consideration for all large-scale assessment projects. The only situation where it may not be useful is one where it is deemed essential to collect common information on all students in a statewide population

---

[6] Editor's note: For information on probability samples, see Womer, _op. cit._

of students. Sampling should be used to maximize the collection of usable information for stated assessment purposes at the lowest possible cost and effort.

* * *

Sampling with total tests is less complicated to administer, but since it is likely to be subject to error in administration and consequently less reliable, in some cases item sampling may be more useful. Therefore, Dr. Womer was asked to prepare an additional statement on the purposes and potential of item sampling. His statement follows.

Item Sampling

The process of item sampling in testing is more useful for one of two purposes:

1. to increase the amount of group test results that can be obtained from students in a given period of time; or
2. to decrease the amount of testing time necessary to obtain large amounts of group test information from students.

For either purpose, it is essential to keep in mind that item sampling is useful for gathering information about groups of students. Thus it is a technique for use with relatively large groups, not a classroom-sized group or even three or four classes within a building.

Example 1

A school system has 500 students in the sixth grade. A standardized reading test is to be administered for a one-shot systemwide survey. The test takes 45 minutes to administer, which is all the time that can be taken from a busy schedule at the end of the year.

Staff are unhappy that only reading is to be surveyed. Some major changes were made in the mathematics curriculum three years before and they feel it would be valuable to survey this subject also. By randomly selecting only 250 of the students to take the reading test, the other 250 could be given a 45-minute mathematics test at the same time.

Example 2

A school system has 1,000 fourth-graders. It is desired to do an in-depth study of student outcomes for 100 different behavioral objectives in mathematics. Each objective requires the use of eight questions. The total of 800 questions would require one student to spend perhaps 15 hours of testing time to attempt all of them.

By randomly dividing up the objectives and items into five different subtests (each with 20 objectives and 160 items), each subtest could be administered to 200 students (randomly selected). This would require only 3 hours of testing time per student (manageable) rather than 15 hours (unmanageable), and group results would still be available for all 100 objectives (800 items).

In either example the results will be usable for group analyses. Any slight reduction in accuracy due to sampling error is apt to be much less than errors due to increasing testing time of students beyond some reasonable amount. Systematic errors due to fatigue, disinterest, poor motivation, teacher concern, and other conditions of testing can easily outweigh a small sampling error.

NATIONAL ASSOCIATION OF SCHOOL COUNSELORS

## STAGES IN DEVELOPING AN EVALUATION PROGRAM

The plan laid out below is no panacea. It is presented only as guidelines from which local associations will want immediately to depart to best serve unique local situations. It is rather more like a check list against which a local may check its present plans and action for aspects it may have slighted or omitted.

### STAGE I

Determine what will be the Association's role in the total evaluation process. Possible ranges of involvement include:

A. Take no Association stand on evaluation. But assure that the grievance machinery is of such a nature and powerful enough that it will respond favorably to Association beliefs and interests in all cases of unfair evaluation.

B. Develop language acceptable to the Association on procedural due process for evaluation for negotiation and/or school district policy.

C. Develop language acceptable to the Association on substantive due process for evaluation for negotiation and/or school district policy.

D. Constitute an Association committee or assign to the local instruction or TEPS committee responsibility for development of a full-blown program for evaluation (to then be negotiated or proposed as school district policy) as delineated in Stages II through XIV.

E. Gain parity representation in an already existing school district committee designated to develop a full-blown program or to accomplish one or more tasks under Stages II through XIV related to evaluation.

F. Take leadership in forming a committee representing a variety of school district constituencies, on which the Association has parity in membership, for the purpose of developing a full-blown evaluation system. Possible constituencies on such a committee, in addition to Association representatives, include: parents, students, board of education, school administration, community pressure groups.

If the decision under Stage I is A, B, or C, the Association will complete the task designated in one of those three and not proceed further.

If the decision under Stage I is D,E, or F, the Association will proceed with Stage II and so on. It is believed that selection of D, E, or F will all require similar activities and a similar progression of stages, although F may require some preliminary political action in bringing into being a working body appropriately representative of several constituencies.

From this point on it is assumed that a group has been constituted or accepted by the Association as suggested in D, E, or F and that all the stages that follow will fully involve this group or its designees.

STAGE II

A. Analyze the evaluation program currently operational in the school district using some agreed-upon criteria for judgment. (The NEA has available 2 levels of devices for this purpose--"Characteristics of a Good Evaluation System" which provides a quick analysis and "An Early Warning Kit on Program and Performance Evaluation" which provides considerably more depth.

B. Determine, in light of the analysis, which of the following tasks are required:

1. Developing a full-blown evaluation program, starting from scratch.
2. Shoring up educational goals and objectives for the schools and relating them to the currently employed evaluation system.
3. Developing training programs for those who apply the evaluation instrument(s).
4. Shoring up the process by which the present instrument(s) is applied.
5. Shoring up due process provisions that relate to evaluation.
6. Minor revision of the present evaluation instrument(s).

7. Balancing performance evaluation with the addition of instruments for program evaluation.

8. Major revision of the present evaluation instrument(s).

9. Development of totally new instrument(s) for performance evaluation.

10. Development of a totally new program evaluation instrument(s).

11. Development of a totally new evaluation program from scratch, including goals and objectives for the schools, goals for evaluation, criteria, who shall evaluate, how data shall be gathered, instrumentation for both program and performance evaluation, training for use of instruments, etc.

> The decisions under Stage II should not be taken lightly. It may not only require analysis of what is on the basis of applying the two NEA tools cited but also some study of the literature on what ought to be in evaluation systems. Having some common understanding among the members of the working group on what is desirable, possible, and what has been accomplished elsewhere can be highly contributory to determining what the local situation requires to bring it up to standard.

Once the decision is made on which of the possibilities in Stage II are to be pursued (several may be selected for pursual concurrently) the stages that follow become relevant or are passed over in light of the selections. If #11 is selected under Stage II, it is expected that in most cases most of the other stages will need to be pursued and generally in the order set down.

## Stage III -- DEVELOPMENT OF OR UP-DATING SCHOOL DISTRICT GOALS AND OBJECTIVES FOR EDUCATION

The following considerations are important in setting goals and objectives for schooling:

A. If the committee constituted or accepted in Stage I is not representative of the community and students, serious consideration should be given to expanding it to reflect these constituencies, at least for accomplishing this stage.

B. Goals and objectives may be updated or developed by one or more of the following procedures:

(a) Reference to state-wide goals and objectives (if available)

1. Reference to ready-made goals and objectives from commercial or non-profit agencies (Phi Delta Kappa has such a list of 14).
2. Use of a ready-made procedure for developing local goals and objectives. A number are available: the Delphi technique, a Phi Delta Kappa gaming procedure, and the Instructional Objectives Exchange are among them.
3. Development of local procedures for determining school district goals and objectives for education.

## STAGE IV -- DEVELOPMENT OF GOALS FOR EVALUATION

Goals for Evaluation should, if at all possible, be tied directly to goals for the schools. At least, evaluation goals should not be inconsistent with the broader school goals.

The following considerations are important in developing goals for evaluation:

A. Determine whether evaluation for improvement of instruction and evaluation for change of status (retention, tenure, promotion) are to be totally separated, and if so, which (or are both) is to be within the purview of the Committee as it carries on its work.

B. Although improvement of instruction and change of status are the 2 major purposes for evaluation, there are others. The Committee should consider all the possibilities and delimit them in terms of what it considers to be of high priority and to be accomplishable in light of resources and time. (A listing of possible goals for evaluation is to be found in "The Evaluation of Educators: A Simulation" produced by the NEA.)

## STAGE V -- DEVELOPMENT OF CRITERIA FOR EVALUATION

This is a most important, sensitive, and difficult task. And it is critical to the activities in Stage VII related to selection, adaptation, or development of instruments for evaluation.

Several key considerations are essential during this stage:

A. An early determination that must be made in this stage is whether both program and performance are to be evaluated. If the decision is to evaluate only performance, the Association should at least be aware of the dangers in agreeing to evaluating teacher performance without concurrently evaluating those program contents, conditions, and resources that make competent performance possible.

B. A second determination is on who shall be evaluated -- teachers, supervisors, administrators, aides and other paraprofessionals and support staff, the Board of Education.

C. A third consideration is whether both professional and personal criteria are to be applied to staff evaluation. That is, will the staff person's deportment in the community, manner of dress, and quality of voice be evaluated along with his ability to individualize instruction, motivate students, and knowledge of subject matter? The key question here is, are effectiveness in promoting student learning to be the only or major criteria on which the staff person is judged or are a variety of ancillary behaviors and characteristics also to be included as criteria.

D. Once determination has been made on the general categories listed in A, B and C, all potential criteria for evaluation need to be considered.

   1. For program evaluation some sources of criteria are NEA's Profiles of Excellence and A Quality Inventory obtainable from Vincent & Olsen, Education Consultants, Salisbury, Connecticut. These may also be used or adapted as evaluation instruments.

## STAGE VI -- DETERMINING WHO SHALL EVALUATE

In some school districts, there may be little leeway in this determination. In other places there is likely wide latitude for decision-making. For example, if the administration has developed and implemented the present evaluation system, and if the Association's position is only to cause the present system and its implementation to perform more fairly and objectively, the matter then is not a question of "who shall evaluate?" But if the Association has assumed one of the positions listed in Stage I requiring that the Association be involved broadly in the evaluation process, then the leeway for decision-making in this phase may be considerable and require the following considerations:

A. Selecting one or more of several possible groups to be directly involved in the evaluation process:

   principals
   other supervisors and administrators
   students
   peers
   self
   parents
   community groups

B. Basing the selection on previous determinations on the goals for evaluation, i.e., improvement of instruction and/or change of status. For example, if the prior decision has been that a goal of evaluation (with Association involvement) is to be change of status, then the use of self-evaluation in implementing the change-of-status goal becomes most difficult.

## STAGE VII -- SELECTING, ADAPTING, DEVELOPING EVALUATION INSTRUMENTS

The tasks to be accomplished in this stage can be relatively simple, or highly complex and of long duration, depending on which of several options are selected in instrumentation. But the size of the task will also depend on the number of criteria selected in Stage V and the decisions on who shall evaluate and who shall be evaluated. For example, if all staff are to be evaluated, as mandated in states like California and Kansas, different instruments may be required for administrators, teachers, aides and the like. And if the decision is for students to be involved in staff evaluation, additional instruments may be required for this purpose.

But the size of the job will be determined more on which of the following alternatives related to instrumentation are selected:

A. Selection of already-developed instruments that are consistent with previously determined goals and criteria for evaluation.

B. Adaptation of already-developed instruments to coincide with previously determined goals (Stage IV) and criteria (Stage V), and to meet other local needs and conditions. This may require obtaining permission to alter or select parts of those commercially-developed instruments that are copyrighted.

C. Development of original instruments that reflect previously-determined goals and criteria for evaluation. Selecting this alternative is likely to require the most time, effort, and talent. It will require not only the ability to develop categories and items that reflect the goals and objectives previously determined, but also some knowledge of item construction, validity and reliability, and other technical matters related to instrumentation. It is likely that the Committee will require outside consultant service if it elects this option.

## STAGE VIII -- TRAINING FOR APPLICATION

It has been said that the results of an evaluation system will be no better than the training of those who apply it. It is obviously near impossible to expect all subjectivity to be eliminated in applying evaluation systems. It is even possible that total objectivity might not be desirable. On the other hand, subjectivity of a high level results in a system's lacking in reliability. Reliability in applying an evaluation system is, to a considerable degree, the consistency with which different evaluators appraise the same behavior, and the consistency with which the same evaluators appraise similar behaviors at different points in time.

Several kinds of activities are appropriate for evaluation training depending on the kinds of instruments that have been selected or developed:

A. Activities for gaining common understanding of instrument terminology.

B. Simulation of observation through films.

C. Paired evaluators in actual classroom observations, followed by discussion and comparison of results.

## STAGE IX -- FIELD TESTING

Any evaluation system to be used for a first time in a new setting, no matter how valid or reliable, should be tried out in a limited number of situations under carefully controlled conditions.

Guidelines need to be developed for the field testing which get at questions like the following:

A. Do the results of its application reflect previously determined goals and criteria? (Are the results valid?)

B. Does it produce results in such a form that they clearly indicate constructive decisions and actions?

C. Are the results reliable?

D. Are the results commensurate with the time spent?

E. Is the application disruptive to the educational process?

F. How do evaluatees react to the application process? (It is wise to survey evaluatees on a wide variety of opinions and attitudes concerning both substance and process of the evaluation system.)

G. What due process provisions are essential?

> Following test and try-out (if possible there should be more than one of these), parts or all of previous stages may need to be repeated before wider application is effected.

## STAGE X -- DEVELOPING DUE PROCESS AGREEMENTS

It should be expected that ideas for due process will have emerged during several earlier stages. But the time of test and try-out is a particularly good opportunity to identify those essential protections that will assure justice to all who are to be evaluated.

Once the essentials are identified, choice will need to be made among three alternatives:

A. Using the general grievance machinery and due process provisions already present in negotiated agreements and school district policy for responding to all matters related to evaluation.

B. Supplementing existing general due process procedures with specific provisions for the evaluation program.

C. Developing a totally independent set of due process procedures for the evaluation program.

> It is unlikely that A alone will be sufficient as an option. Such things as how often evaluation will take place, provision for previous notification to evaluatees, and opportunity for evaluatee response are important considerations that relate specifically to evaluation, and which are not likely to be covered in general grievance machinery provisions.

## STAGE XI -- APPLICATION

If the results of the test and try-out stage have been used to the best advantage, this stage should progress nearly automatically.

Just how involved the Association will become in actual application activities depends on alternatives made in Stage VI. If the Association is to be actively involved, it should be assured that time, resources, and materials are available for carrying on the activities in this stage.

## STAGE XII -- ANALYZING THE RESULTS

The nature of the instruments, the criteria on which they are based, and the original goals selected for evaluation should all be taken into account as the results are analyzed.

If the system is a highly individualized one, analysis of results may be mainly between evaluatee and evaluator. If the system is standardized for groups of staff or, in the case of program instruments, for the total school system rather than by building, general trends in terms of highs, lows, and averages may need to be obtained on a variety of criteria.

Whatever the nature and format of the analysis, its main purpose is to provide data for the constructive completion of the next stage.

## STAGE XIII -- PLANNING AND TAKING ACTIONS BASED ON THE RESULTS

This may be the most important stage of all, and the one demanding the greatest amount of time, creativity, and resources.

If evaluation is to result in the improvement of the instructional program and the improvement of staff performance, no matter at what level they operate, it can be a time-consuming and costly activity, and one which most school districts will need to apply massive effort compared to what is typical in most places today.

Action programs will need to be planned and executed on 2 fronts:

A. To correct deficiencies identified from the program aspects of evaluation.

B. To provide staff in-service education to improve those performance aspects identified as needing upgrading.

The implications of the first for updating curriculum, changing teaching loads, providing time to teach and to plan, for increased materials and improved facilities has important implications for school finance and for the priorities of those who provide financial support to the schools.

And the implications of the second are for massive in-service education programs for all types of school staff--from clerical support personnel to the board of education.

## STAGE XIV -- EVALUATING THE EVALUATION PROGRAM AND RECYCLING

Evaluation should be built in for each stage as the stage gets underway. At this point (assume at the end of year I of full implementation) all applicable stages should be evaluated using the pre-developed, built-in mechanisms, and plans should be made for recycling based on this full evaluation.

## CRITERIA FOR EVALUATING STATE EDUCATION ACCOUNTABILITY SYSTEMS

An acceptable accountability system should respect the complexity of reality. While a good system conceivably could inprove education, a simplistic scheme could deal it deleterious blows and damage the lives of millions of children and teachers. Since children have little defense against ill-conceived schemes, it is incumbent upon professionals to examine such systems seriously.

Education is a serious enterprise. Its essence lies in what happens between a child and his parents, his teachers, and his classmates. These relationships are delicate and susceptible to strong outside influences, and an accountability system must take care not to damage them. Above all, the system must be "livable" for those who are expected to abide by it.

In a pluralistic society an accountability system should promote diversity, not conformity. Opportunities for diversity must exist for the child, the parents, the teacher, the school, and the community. Each entity has a right to be itself. A monolithic system which imposes a single set of values strikes at the very heart of individualism and democratic processes. In short, an accountability system should be responsive to individual differences. Ideally, such a system should strive for personalization.

A final axiom is that an accountability system must be judged on how it will function, not on what it promises. The consequences are far too great for millions of children to rely on vague promises and glib promotion of what the system may do in the future. If the system is simplistic and primitive in its current form, it should not be implemented unless this is done in a carefully controlled laboratory setting where it can be tested and improved, and decisions of consequence to the childrens' lives and teachers' careers should not be based on it.

In addition to these basic principles, there are a number of specific criteria for evaluating state accountability systems.

### STATED PURPOSES AND SPECIFIED USE OF RESULTS

1. The purposes for which the state accountability system is to function should be clear, concise, and understandable to both the profession and the public.

Such statements as "for statewide planning" or "for state decision-making purposes" are too vague to justify the large allocation of resources in time, money, and effort that are being pumped into many state accountability systems.

Stated purposes should describe and provide examples of how education is expected to be improved as a result of implementing the system. Justification on the basis of its contribution to implementing a program planning-budgeting-evaluation system is highly questionable. There is little evidence to date of a positive contribution by PPBES to educational improvement.

2. The uses to which data will be put as they result from the accountability system, should be clearly spelled out in concrete, constructive, and positive terms.

Complete and detailed plans for uses of data resulting from accountability programs should be built in from the beginning. Initiating a program as extensive as is represented in most state accountability systems is a matter of high seriousness and potentially pervasive in its consequences. It should therefore be thought through fully and spelled out clearly and completely before implementation. And the detailing of how the resulting data are to be used should be clearly related to the stated purposes.

## PARTICIPATION AND CONTROL

3. Local control must be retained within a state accountability system.

Besides a long tradition of community control of education in the United States, a strong philosophical and pragmatic case can be made that most educational decisions are best handled close to those who must live with the consequences of those decisions. In education it is the teacher, the parent, and the child himself who are most appreciative of the complexities of the child's world and who are most vitally interested in his personal welfare.

The concrete and personal kinds of information required by the teacher and parent are quite different from the highly abstract and impersonal aggregated data demanded by state bureaucracies. Few would deny that the state has a right to collect data to guide planning, but not at great expense to those at the local level. It would be a grim irony indeed if children's needs and local purposes were to suffer in order to serve the convenience of state administrators.

Just as states have found it necessary to defend many of their rights against the authority of the federal government, so local authorities are justified in defending against state authority their right to be treated as individual entities. Overall, an accountability scheme should not increase the centralization of governmental power.

4. Students, parents, and professionals who will bear the consequences of the accountability system should participate in its development and governance.

Participation means more than being consulted or serving on a committee. It means more than being caught up in an arrangement over which one has no influence. Participation means having influence on decisions and individual recourse to other action when those decisions are disagreeable. It may also ensure greater understanding and commitment.

5. The state accountability system should include explicit provision for holding the state departments of education in general -- and state administrators in particular -- accountalbe to local authorities and professionals.

If accountability can improve the local schools, it can also improve the state departments of education, which are often deficient in the eyes of many educators. Existing state legal responsibility and methods of operation are not sufficient guarantees of accountability any more than is the legal responsibility of the local school board or the tenure status of teachers. A plan should be developed by which the public, students, and professionals can evaluate the competency of the state department in carrying out its charges, preferably in the same manner as local schools and teachers are evaluated.

## DATA COLLECTION

The data collected on the effectiveness of the school must reflect the complexities of the educative process. This would eliminate accountability systems that pursue a few simplistic goals. An English statesman has said:

> As all policy makers know from experience, policy does not consist in prescribing one goal or even one series of goals; but in regulating a system over time in such a way as to optimize the realization of many conflicting relations without wrecking the system in the process. Thus the dominance of technology has infected policy-making with three bogus implications, just admissible in the workshop but lethal in the council chamber. One of these is the habit of accepting goals -- states to be attained once for all -- rather than norms to be held through time, as the typical object of policy. The second is the further reduction of multiple objectives to a single goal, yielding a single criterion of success. The third is the acceptance of effectiveness as the sole criterion by which to choose brtween alternative operations which can be regarded as means to one desired end. The combined effect of these three has been to dehumanize and distort beyond measure the high human function of the government -- that is, regulation -- at all levels.[1]

---

[1] Vickers, Geoffrey. Freedom in a Rocking Boat. Baltimore: Penguin Books 1970.

6. The accountability system should provide for the collection of multiple-outcome data.

No educational program should be evaluated on the basis of a few pieces of information or one or two measures. Assuming that the complex purposes of the whole educational enterprise can be reduced to a few goals, such as teaching "basic skills", or evaluated on a single criterion, no matter what, has catastrophic portent. The more different kinds of data collected, the more likely the evaluation will reflect reality and be fair.

7. The system should provide data for assessing whether program elements and conditions are of a standard of quality to make possible high levels of performance by staff.

Up-to-dateness of curriculum, adequate materials and media, time to plan and to teach, reasonable teaching loads, availability of specialist and clerical services, opportunities for in-service education and decision-making power for teachers -- these and a number of other program conditions and arrangements effect the ability of school staffs to be accountable and must be taken into consideration.

8. The system should provide substantial information on what is going on in the classroom

The classroom atmosphere in which a child spends a significant portion of his life is important, whether or not it results in increased learning. There have been documented instances recently in which children have been beaten to increase achievement scores. This practice is deplorable, even if it is effective. The humanness of the classroom should always be a consideration of high priority.

9. The data collected should include professional judgments.

In the long run there are no substitutes for the reasoned and intuitive judgments of skilled and experienced educators. Social-system measures are too simplistic and primitive to be relied on exclusively. For example, in spite of great publicity, there are only a handful of studies on the efficiency of behavioral objectives, and these studies are equivocal. Ultimately we must rely on the human mind to make judgments. There are no mechanical substitutes.

10. The system should collect data by a variety of techniques from relevant groups and individuals.

Testimonials, interviews, classroom interaction analysis, opinion polls -- all these provide a picture of the richness of educational life and mitigate against making decisions on inadequate and highly abstract information. It is especially important to find out what students are thinking and feeling; they are in the learning setting all the time and their observations are as reliable as any. Parents should also provide judgments on the quality of instruction.

## TESTS

No element of the accountability constellation is more inadequately understood by lay people than standardized achievement tests. The most ill-conceived of all accountability systems would be one which relies heavily on test data gathered at the state level and reported publicly, and which rewards and punishes teachers and administrators on the basis of test results. The potentially destructive effects of such a system are mind-boggling.

11 Under no circumstances should standardized achievement test results be used as the major data in an accountability system.

Tests are not adequate and valid measures of what is taught in school. They are not responsive to school learning unless the teacher teaches the items on the tests. Since tests always sample a domain of behavior, teaching the items on a reading test, for example, does not necessarily mean that students have been taught to read.

The errors involved in testing have been thoroughly explicated by Stake.[2] While some test scores can be useful in diagnosing learning problems and assessing a child's progress, the practitioner also must use many other kinds of information in making decisions, (for example, student interviews, real and simulated performances, products of learning, student self-evaluation, student peer evaluation). When tests are used as a major criterion of learning, their deficiencies are glaring. For the most part they measure only recall-type tasks and shift teaching emphasis from complex mental abilities to those that are simple and easy to measure, this at the expense of long-term retention, relearning ability, and other learning considered by psychologists to be more important.

12. If criterion-referenced tests are substituted for norm-referenced tests in order to overcome some of the testing problems, these tests should be closely scrutinized.

Since no one has been entirely successful in developing criterion-referenced tests, claims for their validity and reliability must be viewed with caution. One major criticism of norm-referencing that criterion-referencing has promised to correct is the lining up of students as related to a measure of central tendency (mean or median) which assures that half will be "below average" no matter how proficient they become in achieving instructional objectives. If movement is to be away from measures of central tendency, these new approaches to validity and reliability will be required.

---

[2] Stake, Robert E. "Measuring What Learners Learn." School Evaluation: The Politics and Process. (Edited by Ernest R. House). Berkeley, Calif: McCutchan Publishing Corp, 1973. pp. 193-223

In addition, averages and means should not be replaced by minimum competency levels, cutting scores, or pass-fail points. This will result only in replacing one statistical device with another for denying opportunities to some students and assuring them to others. It will frustrate the advantage attributed to criterion-referencing of being able to move all students, as rapidly as possible, toward full mastery without the deleterious effects of comparison with others.

13. Test results collected at the state level should not be publicized by school or school district.

Regardless of promises to the contrary, test data collected at the state level are almost invariably made public. The pressures are too great once the data are known to exist. Unless collected anonymously, test results by school and school district should be known only at the local level. In any case, the plan for using the data should be clear in advance.

14. Reports should be tailored for different audiences in order to provide the highest understandability, to avoid misinterpretation, and to assure privacy.

When test scores are reported, the error of measurement should be communicated in understandable form, and only information that can be reported without infringing on the rights of individuals should be included. Information on an individual pupil should not be reported to anyone without the parents' consent. This is consistent with the principle that the district should be accountable to its own constituency rather than to the state agency.

15. If the state desires test data for its own planning purposes, it should use proven matrix sampling techniques which will not reveal schools and which will greatly reduce costs.

Matrix sampling techniques can give an accurate picture of the state by various categories much more efficiently than testing each child with an entire instrument. Otherwise, steps should be taken to protect individual identities. Carefully drawn samples are sufficient for state decision-makers.

16. Districts and schools should not be compared to one another on test scores, nor should a school be judged on the basis of achievement increments or decrements.

Achievement scores are highly influenced by the social and economic conditions within the school district or building attendance area. Therefore, comparisons have little meaning other than to indicate population characteristics. Calculating the gains between two administrations of a test is also highly dubious because of the large and irremediable errors of measurement, turnover in student population, and the like. If such gains are calculated at all, this should be done, not for individual students, but only for large groups in which errors can balance each other out.

17. Rewards or punishments should not be given on the basis of test scores, either group means or individual.

Test scores are highly subject to manipulation by teaching the tests, by selecting the time of year the tests are given, and by controlling conditions and instructions under which they are given. Moreover, tests were never constructed to provide an exact measure of where individuals stand at a given time. Test scores vary so much (called the "error of measurement") that it is possible for individual students to show great gains when no learning or instruction has occurred. The best tests may be a full-year grade equivalent in error. Under a system of rewards and punishments, the temptation to cheat is great indeed. In addition, the OED evaluation of performance contracting has shown that test scores are not raised by rewarding on the basis of results. If a performance contractor is paid on the basis of individual student gains and not penalized for losses, he can make money simply on error fluctuations of test scores.

18. An accountability system should minimize dangerous side effects of relying on test results.

Side effects include suspicion, acrimony, cheating, and a wide assortment of potentially debilitation conditions. An accountability system that leads to wholesale cheating is likely to be counterproductive and result in poorer schools. And, when school administrators and teachers are put under pressure to produce specified results which cannot be assured, there is temptation to teach to the tests or teach the tests. Not only does such pressure mitigate against best professional performance, it encourages unethical acts.

19. If tests are deemed desirable at the local level, the local district should be able to choose among a set of commercially available tests or to develop their own criterion-referenced tests.

Test scores can be useful at the local level, by local option, without comparison with other districts other than what a program supplies. Tests selected or developed to reflect local goals and objectives can serve far more useful purposes than those that attempt to respond broadly to the student population of an entire state.

20. As a general rule, state agencies should not develop their own tests.

State education departments seldom have the manpower or competency to develop their own tests. Although the results of such efforts have not been promising, it would probably be more efficient to use those already available or, if absolutely necessary, to have tests developed on contract.

## COSTS IN DOLLARS, TIME AND PERSONNEL

21. The true cost of the accountability system should be calculated.

Often only a small part of the research and development money necessary to initiate the system is included in costs. The true costs must include the time professionals, children, and others spend providing data. It has been estimated that a complete testing program for a large state, if properly developed and implemented, would cost tens of millions of dollars.

22. The accountability system should not overload professionals or children with providing data.

Many federal programs require the work of a number of local staff to fill out forms, as well as those at the state end of the system to record and analyze the data provided. An accountability system should not be burdonsome to administrators, teachers, or children.

23. The accountability system should require a minimum number of people in the state bureaucracy.

Many state education agency personnel are building careers on the accountability movement. The larger the state bureaucracy becomes in this area, the more it will serve as a lobby to expand continually its own operations past the point of useful returns. The result will be empire building for its own sake.

24. An accountability system should have explicit provisions for the evaluation of its processes and effects.

An accountability program should itself be accountable through a comprehensive plan for auditing its processes, results, and their usefulness for educational improvement.

25. Plans for auditing[3] the success of accountability programs in accomplishing their stated purposes, should be built in as the program is planned.

Auditing programs that are tacked on as afterthoughts or are developed after the program is under way are likely to be ineffective. They appear to suffer inefficiencies, analogous to those in auto air-conditioners installed after the car has left the factory -- they don't respond directly to the nature of the operating mechanism.

15. There should be several kinds of audits applied to a state accountability system.

Certainly the state itself should plan early for its own intensive evaluation of its accountability efforts by a variety of criteria. Foremost among these should be the criterion of how and how much the program has contributed to improving education of the children in the state. In addition, independent outside audits are mandatory. The first line of consideration for such audits should be the teachers of the state. They are the ones who will in the final analysis implement the program and whose professional performance will be influenced most by it. And it is their expertise and professional judgment that should count most. Finally, an outside agency competent in applying the most sophisticated evaluation tools, which is the most independent and impartial and that has the highest credibility to both professionals and the public, should be retained. Such multiple-index and multiple-agency evaluations should take place no less than annually.

---

[3] The term auditing is used here in the context of evaluating the evaluation.